ASYMPTOTIC SIZE AND A PROBLEM WITH SUBSAMPLING AND WITH THE M OUT OF N BOOTSTRAP

BY

DONALD W.K. ANDREWS and PATRIK GUGGENBERGER

COWLES FOUNDATION PAPER NO. 1295

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2010

http://cowles.econ.yale.edu/

Econometric Theory, **26**, 2010, 426–468
doi:10.1017/S0266466609100051

ASYMPTOTIC SIZE AND A PROBLEM WITH SUBSAMPLING AND WITH THE *m* OUT OF *n* BOOTSTRAP

DONALD W.K. ANDREWS
Cowles Foundation, Yale University

PATRIK GUGGENBERGER
University of California at Los Angeles

This paper considers inference based on a test statistic that has a limit distribution that is discontinuous in a parameter. The paper shows that subsampling and m out of n bootstrap tests based on such a test statistic often have asymptotic size—defined as the limit of exact size—that is greater than the nominal level of the tests. This is due to a lack of uniformity in the pointwise asymptotics. We determine precisely the asymptotic size of such tests under a general set of high-level conditions that are relatively easy to verify. The results show that the asymptotic size of subsampling and m out of n bootstrap tests is distorted in some examples but not in others.

1. INTRODUCTION

When the bootstrap is pointwise inconsistent it is common in the literature to consider using subsampling or the m out of n bootstrap instead (see Bretagnolle, 1983; Swanepoel, 1986; Athreya, 1987; Beran and Srivastava, 1987; Shao and Wu, 1989; Wu, 1990; Eaton and Tyler, 1991; Politis and Romano, 1994; Shao, 1994, 1996; Beran, 1997; Bickel, Götze, and van Zwet, 1997; Andrews, 1999, 2000; Politis, Romano, and Wolf, 1999 (hereafter PRW); Romano and Wolf, 2001; Guggenberger and Wolf, 2004; and Lehmann and Romano, 2005). (Here n denotes the sample size, b denotes the subsample size, and m denotes the bootstrap sample size.) Minimal conditions are needed for subsampling and m out of n bootstrap tests and confidence intervals (CIs) to have asymptotically correct rejection rates and coverage probabilities under pointwise asymptotics, e.g., see Politis and Romano (1994) and PRW. Given these results, subsampling and the

This paper previously circulated under the title "The Limit of Finite-Sample Size and a Problem with Subsampling" as Cowles Foundation Discussion Paper No. 1605. Andrews was supported by NSF grants SES-0417911 and SES-0751517. Guggenberger was supported by a Sloan fellowship, a UCLA faculty research grant in 2005 and NSF grant SES-0748922. For helpful comments, we thank two referees, the co-editor Richard Smith, Victor Chernozhukov, In Choi, Russell Davidson, Hannes Leeb, David Pollard, Azeem Shaikh, Jim Stock, Michael Wolf, and the participants at various seminars and conferences at which the paper was presented. Address correspondence to Donald W.K. Andrews, Department of Economics, Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, Yale Station, New Haven, CT 06520-8281; e-mail: donald.andrews@yale.edu.

 © Cambridge University Press, 2009 0266-4666/10 $15.00

m out of n bootstrap have been viewed in the literature as cure-all methods that are asymptotically valid under very weak assumptions.

This paper considers subsampling and the m out of n bootstrap in a broad class of problems in econometrics and statistics in which a test statistic has a discontinuity in its asymptotic distribution as a function of the true distribution that generates the data. These are precisely the sorts of scenarios where one may want to employ subsampling or the m out of n bootstrap. We show that subsampling and the m out of n bootstrap are not necessarily asymptotically valid in a uniform sense in these problems. Specifically, the asymptotic sizes, i.e., the limits of the exact (or finite-sample) sizes, of subsampling and m out of n bootstrap tests can exceed their nominal level—in some cases by a lot.[1,2] This is a serious problem because it implies that in such cases the size of the test is far from its nominal level even for large n. Correct size is the standard measure of validity of a test in finite samples. Asymptotic size is a large sample approximation to exact size. Our results determine when a problem concerning asymptotic size arises, when it does not, and the magnitude of the problem. The latter is provided by an explicit formula for asymptotic size, which is the main contribution of this paper. The reason for distorted asymptotic size, when it occurs, is a lack of uniformity in the asymptotics.

We now briefly illustrate the problem with subsampling by considering a simple boundary example. Suppose $X_i \sim$ i.i.d. $N(\theta_0, 1)$ for $i = 1, \ldots, n$ and $\theta_0 \geq 0$. The maximum likelihood estimator of θ_0 is $\widehat{\theta}_n = \max\{\overline{X}_n, 0\}$, where $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$. The distribution of the normalized estimator, T_n, is

$$T_n = n^{1/2}(\widehat{\theta}_n - \theta_0) = \max\{n^{1/2}(\overline{X}_n - \theta_0), -n^{1/2}\theta_0\} \sim \max\{Z, -h\}, \tag{1}$$

where $Z \sim N(0, 1)$ and $h = n^{1/2}\theta_0$. The jth subsample estimator based on a subsample of size $b = o(n)$ starting at the jth observation is $\widehat{\theta}_{b,j} = \max\{\overline{X}_{b,j}, 0\}$, where $\overline{X}_{b,j} = b^{-1} \sum_{i=j}^{j+b-1} X_i$. The distribution of the normalized subsample estimator, $T_{b,j}$, is

$$\begin{aligned} T_{b,j} = b^{1/2}(\widehat{\theta}_{b,j} - \theta_0) &= \max\{b^{1/2}(\overline{X}_{b,j} - \theta_0), -b^{1/2}\theta_0\} \\ &\sim \max\{Z, -(b/n)^{1/2} h\}. \end{aligned} \tag{2}$$

The distributions in (1) and (2) can be quite different. For example, when h is large and $(b/n)^{1/2} h$ is small, the distribution in (1) is approximately that of Z while that in (2) is approximately that of $\max\{Z, 0\}$. Clearly, in such cases, the subsampling distribution gives a very poor approximation of the full-sample distribution in the left tail. Hence, a lower one-sided subsampling confidence interval (C.I) for θ_0 performs poorly. Furthermore, equal-tailed and symmetric two-sided subsampling confidence intervals also perform poorly. Note that these are finite-sample results. The asymptotic sizes of these confidence intervals detect the problems. With nominal level .95, the asymptotic sizes for lower, equal-tailed, and symmetric CIs are found to be .50, .475, and .90, respectively.

In spite of the clear finite-sample problems with subsampling in this example, pointwise asymptotics fail to detect any problem. The reason is that for θ_0 fixed at zero, one has $h = n^{1/2}\theta_0 = 0$ in (1) and $(b/n)^{1/2}h = b^{1/2}\theta_0 = 0$ in (2) for all n. For θ_0 fixed at a positive value, one has $h = n^{1/2}\theta_0 \to \infty$ in (1) and $(b/n)^{1/2}h = b^{1/2}\theta_0 \to \infty$ in (2) as $n \to \infty$. These asymptotic results, however, do not hold uniformly. This is seen by considering the sequence of parameter values $\theta_0 = h/n^{1/2}$, where $h \neq 0$ is fixed, which approaches zero as $n \to \infty$. For such a sequence, the difference between the asymptotic distributions of the full-sample and subsample statistics becomes apparent. In consequence, the asymptotic size exceeds the nominal size, which reflects the finite-sample situation. (We use this CI example because of its simplicity. The results of the paper concern tests. General results for CIs are given in Andrews and Guggenberger (2009b).)

More generally, the idea behind problems with subsampling is as follows. Suppose (i) one is interested in testing $H_0 : \theta = \theta_0$, (ii) a nuisance parameter $\gamma \in R$ appears under H_0, (iii) we have a test statistic T_n for which large values lead to rejection of H_0, and (iv) the asymptotic distribution of T_n when $\theta = \theta_0$ is discontinuous at $\gamma = 0$. Typically in such situations the asymptotic distribution of T_n under $\theta = \theta_0$ and under any drifting sequence of parameters $\{\gamma_n = (h + o(1))/n^r : n \geq 1\}$ depends on the "localization parameter" $h \in R$. That is,

$$T_n \to_d J_h \quad \text{as} \quad n \to \infty \quad \text{under} \quad \theta = \theta_0 \quad \text{and} \quad \{\gamma_n = (h + o(1))/n^r : n \geq 1\}, \tag{3}$$

where J_h is some distribution. (The constant $r > 0$ is the smallest constant such that the distribution of T_n under γ_n is contiguous to its distribution under $\gamma = 0$. Usually $r = 1/2$, but not always, for example, in an autoregressive model with a discontinuity at a unit root, $r = 1$.)

We assume that the subsample size b satisfies $b \to \infty$ and $b/n \to 0$. By (3),

$$T_b \to_d J_0 \quad \text{under} \quad \theta = \theta_0 \quad \text{and} \quad \{\gamma_n = (h + o(1))/n^r : n \geq 1\} \tag{4}$$

because $\gamma_n = (h + o(1))/n^r = (b/n)^r (h + o(1))/b^r = o(1)/b^r$. Subsample statistics with subsample size b have the same asymptotic distribution J_0 as T_b has. In consequence, the subsampling critical value converges in probability to $c_0(1-\alpha)$, the $1-\alpha$ quantile of J_0. In contrast, T_n converges in distribution to J_h and T_n requires the critical value $c_h(1-\alpha)$, the $1-\alpha$ quantile of J_h, in order to have an asymptotic null rejection probability of α under $\{\gamma_n : n \geq 1\}$. If $c_0(1-\alpha) < c_h(1-\alpha)$, the subsampling test over-rejects asymptotically under $\{\gamma_n : n \geq 1\}$. This implies that it has asymptotic size greater than α. On the other hand, if $c_0(1-\alpha) > c_h(1-\alpha)$, then the subsampling test under-rejects asymptotically and is asymptotically nonsimilar in a uniform sense.

There are other (noncontiguous) sequences of drifting parameters that can cause problems for subsampling. Suppose $\gamma_n = (g + o(1))/b^r$ for $g \in R$. Then, by (3), T_b has asymptotic distribution J_g and the subsampling critical value converges in probability to $c_g(1-\alpha)$. On the other hand, $\gamma_n = (n/b)^r(g + o(1))/n^r$ and

$(n/b)^r \to \infty$, so the full-sample statistic T_n converges to J_∞ (when $g \neq 0$), which is assumed to be the asymptotic distribution of T_n when γ_n is farther from 0 than $O(n^{-r})$ (i.e., when the distribution of T_n under γ_n is not contiguous to its distribution under $\gamma = 0$). If $c_g(1-\alpha) < c_\infty(1-\alpha)$, where $c_\infty(1-\alpha)$ denotes the $1-\alpha$ quantile of J_∞, the subsampling test over-rejects asymptotically under $\{\gamma_n : n \geq 1\}$.

As stated above, we are interested in the asymptotic size of a test because it can be used to approximate the exact size of the test. In this paper, we show that sequences of the two types discussed above determine the asymptotic size of the subsampling test. Specifically, the asymptotic size equals the maximum of $1 - J_h(c_g(1-\alpha))$ over those pairs $(g,h) \in (R \cup \{\infty\})^2$ such that $g = 0$ if $h < \infty$ and $g \in R \cup \{\infty\}$ if $h = \infty$.

In many models, the test statistic T_n depends on more than just a scalar nuisance parameter γ. For example, in some cases T_n depends on two nuisance parameters (γ_1, γ_2) and its asymptotic distribution is discontinuous in γ_1, depends on γ_2, but is not discontinuous in γ_2. In such cases, the asymptotic distribution of T_n depends on a localization parameter h_1 analogous to h above and the fixed value of γ_2. The asymptotic behavior of subsampling tests in this case is as described above with h_1 in place of h except that the conditions for the asymptotic size to be α or less must hold for each value of γ_2. The results given below allow for cases of this type, including cases where γ_1 and γ_2 are vectors. The results given below also allow for the common case where a third nuisance parameter γ_3 appears and has the property that it does not affect the asymptotic distribution of T_n. For example, γ_3 may be an infinite-dimensional parameter such as the distribution of an error term that is normalized to have mean zero and variance one.

The paper gives asymptotic results for subsampling tests with subsample size b. Such results also apply to m out of n bootstrap tests with bootstrap size $m = b$ when the observations are i.i.d. and $b^2/n \to 0$. (This holds because the difference between sampling with and without replacement goes to zero as $n \to \infty$ in this case, see PRW, p. 48.) In consequence, in the remainder of the paper, we focus on subsampling procedures only.

The results of the paper are shown below, in Andrews and Guggenberger (2005, 2009a, 2009b, 2009c), and in Guggenberger (2007, 2009) to apply to a wide variety of examples. In these examples, the asymptotic sizes of subsampling tests and CIs are found to vary widely depending on the particular model and statistic considered and on the type of inference considered, e.g., upper or lower one-sided or symmetric or equal-tailed two-sided tests or CIs.

In this paper, the general results are used to show the following. (i) In a model with a nuisance parameter near a boundary, lower one-sided, upper one-sided, symmetric two-sided, and equal-tailed two-sided subsampling tests with nominal level .05 have asymptotic sizes of (approximately) .50, .50, .10, and .525, respectively. (ii) In an instrumental variables (IVs) regression model with potentially weak IVs, all nominal level $1-\alpha$ one-sided and two-sided subsampling tests concerning the coefficient on an exogenous variable and based on the two-stage least squares (2SLS) estimator have asymptotic size equal to one (for both partially- and fully-studentized test statistics).

Results established elsewhere using the approach of this paper are as follows: (iii) In an autoregressive model with an intercept and an autoregressive root that may be near unity, as considered in Romano and Wolf (2001), equal-tailed and symmetric two-sided subsampling CIs of nominal level .95 based on least squares estimators are found to have asymptotic sizes of (approximately) .60 and .95, respectively, see Andrews and Guggenberger (2009a). When conditional heteroskedasticity is present, the same asymptotic sizes are obtained for subsampling procedures based on generalized least squares estimators and heteroskedasticity-consistent standard error estimators. (iv) In models where (partially-identified) parameters are restricted by moment inequalities, subsampling tests and CIs based on suitable test statistics have correct asymptotic size, see Andrews and Guggenberger (2009b). (v) A subsampling CI of nominal level $1-\alpha$ based on a post-consistent-model-selection estimator (such as one based on BIC), a shrinkage estimator, or a super-efficient estimator is found to have asymptotic size of zero, see Andrews and Guggenberger (2009c).

The results of the paper also are shown in Andrews and Guggenberger (2005, 2009a) to apply to inference for (vi) post-conservative model-selection procedures (such as those based on AIC), (vii) models with lack of identification at some point(s) in the parameter space, such as models with weak instruments where the focus is on tests concerning the coefficient of an endogenous variable, and (viii) parameters of interest that may be near a boundary. In all of the examples listed above except (iv), some types of subsampling procedures (such as equal-tailed ones) do not have asymptotic size equal to their nominal level, although other types sometimes do. The results of the paper can be applied with some modifications to inference in linear instrumental variables models when the instruments are locally nonexogenous; see Guggenberger (2009).

The results of the paper also can be applied to inference for (ix) predictive regression models with nearly-integrated regressors, (x) threshold autoregressive models, (xi) tests of stochastic dominance, (xii) nondifferentiable functions of parameters, (xiii) differentiable functions of parameters that have zero first-order derivatives, and (xiv) tests for a breakpoint with small magnitude.

The testing results of the paper extend to CIs with some adjustments, see Andrews and Guggenberger (2009b). Adjustments are needed because a CI for θ requires uniformity over the nuisance parameters *and* the parameter of interest θ, whereas a test concerning θ only requires uniformity over the nuisance parameters (because θ is fixed by the null hypothesis).

Problems arising from lack of uniformity in asymptotics have long been recognized in the statistical literature. For example, see LeCam (1953), Bahadur and Savage (1956), Rao (1963, 1973), Hájek (1971), Pfanzagl (1973), and Sen (1979). More recent references include Loh (1985), Gleser and Hwang (1987), Sen and Saleh (1987), Stock (1991), Cavanagh, Elliot, and Stock (1995), Hall and Jing (1995), Kabaila (1995), Dufour (1997), Staiger and Stock (1997), Pötscher

(2002), Anatolyev (2004), Imbens and Manski (2004), and Leeb and Pötscher (2005, 2006). For the bootstrap, different types of uniformity are discussed in, e.g., Bickel and Freedman (1981), Beran (1984), Romano (1989), Giné and Zinn (1990), and Sheehy and Wellner (1992).

In the specific context of subsampling and the m out of n bootstrap, however, the only other papers in the literature that we are aware of that raise the issue of uniformity in the sense discussed in this paper are as follows: (i) Dümbgen (1993) shows that under drifting sequences of true parameters the asymptotic distribution of the m out of n bootstrap estimator of the distribution of a nondifferentiable function of a statistic need not equal the asymptotic distribution of the nondifferentiable function of the full-sample statistic. The latter property indicates that caution is warranted. However, this property does not imply that an m out of n bootstrap test or CI necessarily has incorrect asymptotic size. Examples where it holds, but subsampling and m out of n bootstrap tests and CIs have correct asymptotic size, include symmetric two-sided CIs in an autoregressive model with a possible unit root, see Andrews and Guggenberger (2009a), and tests and CIs in the moment inequality model, see Andrews and Guggenberger (2009b).

(ii) Beran (1997, p.15) notes that the pointwise m out of n bootstrap convergence typically is not locally uniform at parameter points that are not locally asymptotically equivariant, but does not discuss the consequences. In particular, results in Beran (1997) do not show that an m out of n bootstrap or subsampling test or CI has incorrect asymptotic size.

(iii) Andrews (2000, p.403) notes that subsampling is not consistent for the distribution of an estimator when the true parameter converges to a boundary at rate $1/n^{1/2}$.

(iv) Samworth (2003) shows by simulation that the m out of n bootstrap can perform poorly when estimating the distribution of Hodges estimator, but does not provide any asymptotic results.

(v) Romano and Shaikh (2005, 2008) provide high-level sufficient conditions for uniform validity of subsampling, but do not discuss invalidity in any contexts. Their conditions have been verified in the context of inference based on moment inequalities for a simple test statistic that is not scale equivariant, but to the best of our knowledge have not been verified in any other models.

(vi) Mikusheva (2007) presents a counterexample to the uniform asymptotic validity of an equal-tailed subsampling CI in the context of an autoregressive model.

Romano and Shaikh (2005, 2008) and Mikusheva (2007) were written independently of and at about the same time as the present paper. As far as we are aware, the first explicit counterexample to the uniform asymptotic validity of a subsampling or m out of n bootstrap test or CI was given by the present authors for a boundary example in early work on this paper in November 2004.

The remainder of the paper is organized as follows. Section 2 describes the basic testing setup. Sections 3 and 4 specify the general assumptions and asymptotic results of the paper for one-sided and symmetric two-sided tests. Section 5 extends the results to equal-tailed two-sided tests. Section 6 discusses two examples of the general results. Section 7 gives sufficient conditions for a technical assumption used in the paper. Section 8 provides proofs.

2. TESTING SETUP

We now describe the general testing setup. We are interested in tests concerning a parameter $\theta \in R^d$ in the presence of a nuisance parameter $\gamma \in \Gamma$, where Γ is specified below. The null hypothesis is $H_0 : \theta = \theta_0$ for some $\theta_0 \in R^d$. The alternative hypothesis may be one-sided or multi-sided. Let $T_n(\theta_0)$ denote a real-valued test statistic for testing H_0 based on a sample of size n. The leading case is when $T_n(\theta_0)$ is a t statistic, but the results cover other test statistics. The focus of this paper is on the behavior of tests when the asymptotic null distribution of $T_n(\theta_0)$ depends on the nuisance parameter γ and is discontinuous at some value(s) of γ.

A test rejects the null hypothesis when $T_n(\theta_0)$ exceeds some critical value. We consider two types of critical values for use with the test statistic $T_n(\theta_0)$. The first is a *fixed critical value* (FCV) and is denoted $c_{Fix}(1-\alpha)$, where $\alpha \in (0,1)$ is the nominal size of the FCV test. The FCV test rejects H_0 when $T_n(\theta_0) > c_{Fix}(1-\alpha)$. A common choice when $T_n(\theta_0)$ has the same asymptotic distribution for all fixed γ that are not points of discontinuity is $c_{Fix}(1-\alpha) = c_\infty(1-\alpha)$, where $c_\infty(1-\alpha)$ denotes the $1-\alpha$ quantile of J_∞ and J_∞ is the asymptotic null distribution of $T_n(\theta_0)$ when γ is fixed and is not a point of discontinuity.

The second type of critical value that we consider is a subsampling critical value. Let b denote the subsample size, which depends on n. The number of different subsamples of size b is q_n. With i.i.d. observations, there are $q_n = n!/((n-b)!b!)$ different subsamples of size b. With time series observations, there are $q_n = n-b+1$ subsamples, each consisting of b consecutive observations. The subsample statistics that are used to construct the subsampling critical value are denoted by $\{\widehat{T}_{n,b,j} : j = 1,\ldots,q_n\}$.

Let $\{T_{n,b,j}(\theta_0) : j = 1,\ldots,q_n\}$ be subsample statistics that are defined exactly as $T_n(\theta_0)$ is defined, but are based on subsamples of size b rather than the full sample. The empirical distribution of $\{T_{n,b,j}(\theta_0) : j = 1,\ldots,q_n\}$ is

$$U_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta_0) \le x). \tag{5}$$

In most cases, the subsample statistics $\{\widehat{T}_{n,b,j} : j = 1,\ldots,q_n\}$ are defined to satisfy one or other of the following assumptions.

Assumption Sub1. $\widehat{T}_{n,b,j} = T_{n,b,j}(\widehat{\theta}_n)$ for all $j \le q_n$, where $\widehat{\theta}_n$ is an estimator of θ.

Assumption Sub2. $\widehat{T}_{n,b,j} = T_{n,b,j}(\theta_0)$ for all $j \le q_n$.

The estimator $\widehat{\theta}_n$ in Assumption Sub1 usually is chosen to be an estimator that is consistent under both the null and alternative hypotheses.

Let $L_{n,b}(x)$ and $c_{n,b}(1-\alpha)$ denote the empirical distribution function and $1-\alpha$ sample quantile, respectively, of the subsample statistics $\{\widehat{T}_{n,b,j} : j = 1,\ldots,q_n\}$. By definition,

$$L_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(\widehat{T}_{n,b,j} \le x) \quad \text{for } x \in R \quad \text{and}$$

$$c_{n,b}(1-\alpha) = \inf\{x \in R : L_{n,b}(x) \ge 1-\alpha\}. \tag{6}$$

The subsampling test rejects $H_0 : \theta = \theta_0$ if

$$T_n(\theta_0) > c_{n,b}(1-\alpha). \tag{7}$$

The exact and asymptotic sizes of FCV and subsampling tests are

$$ExSz_n(\theta_0) = \sup_{\gamma \in \Gamma} P_{\theta_0,\gamma}(T_n(\theta_0) > c_{1-\alpha}) \quad \text{and}$$

$$AsySz(\theta_0) = \limsup_{n\to\infty} ExSz_n(\theta_0), \tag{8}$$

where $c_{1-\alpha} = c_{Fix}(1-\alpha)$ or $c_{1-\alpha} = c_{n,b}(1-\alpha)$ and $P_{\theta,\gamma}(\cdot)$ denotes probability when the true parameters are (θ, γ). Uniformity over $\gamma \in \Gamma$, which is built into the definition of $AsySz(\theta_0)$, is necessary for the asymptotic size to give a good approximation to the finite-sample size. Obviously, the specification of the parameter space Γ plays a key role in the exact size of a test. We are interested here in problems in which the elements of Γ at which discontinuities of the asymptotic distribution of the test statistic occur are parameter values that are empirically relevant. In addition to asymptotic size, we also are interested in the minimum rejection probability of the test and its limit: $MinRP_n(\theta_0) = \inf_{\gamma \in \Gamma} P_{\theta_0,\gamma}(T_n(\theta_0) > c_{1-\alpha})$ and $AsyMinRP(\theta_0) = \liminf_{n\to\infty} MinRP_n(\theta_0)$. If $\alpha - AsyMinRP(\theta_0) > 0$, then the test is not asymptotically similar in a uniform sense and, hence, may sacrifice power.

We now introduce a running example that is used for illustrative purposes.

Example 1

We consider a testing problem where a nuisance parameter may be near a boundary of the parameter space under the null hypothesis. Suppose $\{X_i \in R^2 : i \le n\}$ are i.i.d. with distribution F. Then

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}, \quad E_F X_i = \begin{pmatrix} \theta \\ \mu \end{pmatrix}, \quad \text{and} \quad \mathrm{Var}_F(X_i) = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}. \tag{9}$$

The null hypothesis is $H_0 : \theta = 0$, i.e., $\theta_0 = 0$. (The results below are invariant to the choice of θ_0.) The parameter space for the nuisance parameter μ is $[0, \infty)$.

We consider lower and upper one-sided tests and symmetric and equal-tailed two-sided tests of nominal level α. Each test is based on a studentized test statistic $T_n(\theta_0)$, where $T_n(\theta_0) = T_n^*(\theta_0), -T_n^*(\theta_0)$, or $|T_n^*(\theta_0)|$, and $T_n^*(\theta_0) = n^{1/2}(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_{n1}$.

The estimators $(\widehat{\theta}_n, \widehat{\sigma}_{n1})$ of (θ, σ_1) are defined as follows. Let $\widehat{\sigma}_{n1}$, $\widehat{\sigma}_{n2}$, and $\widehat{\rho}_n$ denote any consistent estimators of σ_1, σ_2, and ρ. We suppose that $\widehat{\sigma}_{n1}$ is scale equivariant, i.e., the distribution of $\widehat{\sigma}_{n1}/\sigma_1$ does not depend on σ_1, as is true of most estimators of σ_1. Let $(\widehat{\theta}_n, \widehat{\mu}_n)$ be the Gaussian quasi-maximum likelihood (ML) estimator of (θ, μ) under the restriction that $\mu \geq 0$ and under the assumption that the standard deviations and correlation of X_{i1} and X_{i2} equal $\widehat{\sigma}_{n1}$, $\widehat{\sigma}_{n2}$, and $\widehat{\rho}_n$. This allows for the case where $(\widehat{\theta}_n, \widehat{\mu}_n, \widehat{\sigma}_{n1}, \widehat{\sigma}_{n2}, \widehat{\rho}_n)$ is the Gaussian quasi-ML estimator of $(\theta, \mu, \sigma_1, \sigma_2, \rho)$ under the restriction $\mu \geq 0$. Alternatively, $\widehat{\sigma}_{n1}$, $\widehat{\sigma}_{n2}$, and $\widehat{\rho}_n$ could be the sample standard deviations and correlation of X_{i1} and X_{i2}. A Kuhn-Tucker maximization shows that

$$\widehat{\theta}_n = \overline{X}_{n1} - (\widehat{\rho}_n \widehat{\sigma}_{n1}) \min(0, \overline{X}_{n2}/\widehat{\sigma}_{n2}), \quad \text{where}$$
$$\overline{X}_{nj} = n^{-1} \sum_{i=1}^{n} X_{ij} \quad \text{for} \quad j = 1, 2. \tag{10}$$

The FCVs employed in this example are the usual standard normal critical values that ignore the fact that μ may be on or near the boundary. They are $z_{1-\alpha}$, $z_{1-\alpha}$, and $z_{1-\alpha/2}$, respectively, for the upper, lower, and symmetric versions of the test. The subsampling critical values are given by $c_{n,b}(1-\alpha)$ obtained from the subsample statistics $\{T_{n,b,j}(\widehat{\theta}_n) : j \leq q_n\}$ that satisfy Assumption Sub1. (The same results as given below hold under Assumption Sub2.)

3. ASSUMPTIONS

3.1. Parameter Space

The model is indexed by a parameter γ that has up to three components: $\gamma = (\gamma_1, \gamma_2, \gamma_3)$. The points of discontinuity of the asymptotic distribution of the test statistic of interest are determined by the first component, $\gamma_1 \in R^p$. We assume that the discontinuities occur when one or more elements of γ_1 equal zero. The parameter space for γ_1 is $\Gamma_1 \subset R^p$. The second component of γ, γ_2 ($\in R^q$), also affects the limit distribution of the test statistic, but does not affect the distance of the parameter γ to the point of discontinuity. The parameter space for γ_2 is $\Gamma_2 \subset R^q$. The third component of γ, γ_3, does not affect the limit distribution of the test statistic. It is assumed to be an element of an arbitrary space $\mathcal{T}_3$. Infinite dimensional γ_3 parameters, such as error distributions, arise frequently in examples. Due to the central limit theorem (CLT), the asymptotic distribution of a test statistic often does not depend on the specific error distribution—only on whether it has certain moments finite and uniformly bounded (and for nonstudentized statistics on its scale). The parameter space for γ_3 is $\Gamma_3(\gamma_1, \gamma_2)$ ($\subset \mathcal{T}_3$), which may depend on γ_1 and γ_2.

The parameter space for γ is

$$\Gamma = \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\}. \tag{11}$$

Let $\lfloor$ denote the left endpoint of an interval that may be open or closed at the left end. Define $\rfloor$ analogously for the right endpoint.

Assumption A. (i) Γ satisfies (11) and (ii) $\Gamma_1 = \prod_{m=1}^{p} \Gamma_{1,m}$, where $\Gamma_{1,m} = \lfloor \gamma_{1,m}^{\ell}, \gamma_{1,m}^{u} \rfloor$ for some $-\infty \leq \gamma_{1,m}^{\ell} < \gamma_{1,m}^{u} \leq \infty$ that satisfy $\gamma_{1,m}^{\ell} \leq 0 \leq \gamma_{1,m}^{u}$ for $m = 1, \ldots, p$.

Assumption A(ii) is satisfied in many examples, including all of those considered in Andrews and Guggenberger (2005, 2009a, 2009b, 2009c) except the moment inequality model when restrictions arise such that one inequality cannot hold as an equality if another inequality holds as an equality. The results in Andrews and Guggenberger (2009b) do not require Assumption A(ii). Neither do the results in Romano and Shaikh (2008) (but the latter do not cover any cases in which subsampling does not have correct asymptotic size).

In the "continuous limit" case, in which no discontinuity of the asymptotic distribution of the test statistic occurs, no parameter γ_1 appears and $p = 0$.

Example 1 (cont.)

In this example, the vector of nuisance parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ is defined by $\gamma_1 = \mu/\sigma_2$, $\gamma_2 = \rho$, and $\gamma_3 = (\sigma_1, \sigma_2, F)$. In Assumption A, set $\Gamma_1 = R_+$, where $R_+ = \{x \in R : x \geq 0\}$, $\Gamma_2 = [-1, 1]$, and

$$\begin{aligned}\Gamma_3(\gamma_1, \gamma_2) = \{(\sigma_1, \sigma_2, F) : \sigma_1 \in (0, \infty),\ \sigma_2 \in (0, \infty), \mathrm{E}_F||X_i||^{2+\delta} \leq M,\\ \mathrm{E}_F X_i = (0, \mu)', \mathrm{Var}_F(X_{i1}) = \sigma_1^2,\ \mathrm{Var}_F(X_{i2}) = \sigma_2^2,\\ Corr_F(X_{i1}, X_{i2}) = \gamma_2,\ \&\ \gamma_1 = \mu/\sigma_2\}\end{aligned} \tag{12}$$

for some $M < \infty$ and $\delta > 0$. Given these definitions, Assumption A holds. The condition $\mathrm{E}_F||X_i||^{2+\delta} \leq M$ in $\Gamma_3(\gamma_1, \gamma_2)$ ensures that the Liapunov CLT applies in (15)–(17) below. In $\Gamma_3(\gamma_1, \gamma_2)$, $\mathrm{E}_F X_{i1} = 0$ because the results given are under the null hypothesis.

The null distribution of $T_n^*(\theta_0)$ is invariant to σ_1^2 because $\widehat{\sigma}_{n1}$ is scale equivariant. Hence, for simplicity and without loss of generality, when analyzing the asymptotic properties of the tests in this example, we assume that $\sigma_1^2 = 1$ for all n and $\Gamma_3(\gamma_1, \gamma_2)$ is restricted correspondingly.

3.2. Convergence Assumptions

In this section, the true value of θ is the null value θ_0 and all limits are as $n \to \infty$. For a sequence of constants $\{\kappa_n : n \geq 1\}$, let $\kappa_n \to [\kappa_{1,\infty}, \kappa_{2,\infty}]$ denote that $\kappa_{1,\infty} \leq \liminf_{n\to\infty} \kappa_n \leq \limsup_{n\to\infty} \kappa_n \leq \kappa_{2,\infty}$. For an arbitrary distribution G, let $G(\cdot)$ denote the distribution function (df) of G, let $G(x-)$

denote the limit from the left of $G(\cdot)$ at x, and let $C(G)$ denote the set of continuity points of $G(\cdot)$. Define the $1-\alpha$ quantile, $q(1-\alpha)$, of a distribution G by $q(1-\alpha) = \inf\{x \in R : G(x) \geq 1-\alpha\}$. The distribution J_h considered below is the distribution of a proper random variable that is finite with probability one. Let $R_+ = \{x \in R : x \geq 0\}$, $R_- = \{x \in R : x \leq 0\}$, $R_{+,\infty} = R_+ \cup \{\infty\}$, $R_{-,\infty} = R_- \cup \{-\infty\}$, $R_\infty = R \cup \{\pm\infty\}$, $R_+^p = R_+ \times \ldots \times R_+$ (with p copies), and $R_\infty^p = R_\infty \times \ldots \times R_\infty$ (with p copies).

Let $r > 0$ denote a *rate of convergence index* such that when the true parameter γ_1 satisfies $n^r \gamma_1 \to h_1$, then the test statistic $T_n(\theta_0)$ has an asymptotic distribution that depends on the localization parameter h_1 (see Assumption B below). The constant r is the smallest constant such that sequences of parameters γ_1 of order $O(1/n^r)$ yield distributions of the observations that are contiguous to the distributions at $\gamma_1 = 0_p$, which is the discontinuity point of the asymptotic distribution of $T_n(\theta_0)$. In most examples, $r = 1/2$, but in the unit root example considered in Andrews and Guggenberger (2009a), $r = 1$.

We now define the index set for the different asymptotic null distributions of the test statistic $T_n(\theta_0)$ of interest. Let

$$\begin{aligned} H = \{ & h = (h_1, h_2) \in R_\infty^{p+q} : \exists \{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\} \\ & \text{such that} \quad n^r \gamma_{n,1} \to h_1 \quad \text{and} \quad \gamma_{n,2} \to h_2\}. \end{aligned} \tag{13}$$

For notational simplicity, we write $h = (h_1, h_2)$, rather than $(h_1', h_2')'$, even though h is a $p+q$ column vector. Under Assumption A, it follows that

$$H = H_1 \times H_2,\ H_1 = \prod_{m=1}^{p} \begin{cases} R_{+,\infty} & \text{if } \gamma_{1,m}^\ell = 0 \\ R_{-,\infty} & \text{if } \gamma_{1,m}^u = 0 \\ R_\infty & \text{if } \gamma_{1,m}^\ell < 0 \text{ and } \gamma_{1,m}^u > 0, \end{cases} \qquad H_2 = \text{cl}(\Gamma_2), \tag{14}$$

where $\text{cl}(\Gamma_2)$ is the closure of Γ_2 with respect to R_∞^q. For example, if $p = 1$, $\gamma_{1,1}^\ell = 0$, and $\Gamma_2 = R^q$, then $H_1 = R_{+,\infty}$, $H_2 = R_\infty^q$, and $H = R_{+,\infty} \times R_\infty^q$.

DEFINITION OF $\{\gamma_{n,h} : n \geq 1\}$. *Given $r > 0$ and $h = (h_1, h_2) \in H$, let $\{\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3}) : n \geq 1\}$ denote a sequence of parameters in Γ for which $n^r \gamma_{n,h,1} \to h_1$ and $\gamma_{n,h,2} \to h_2$.*

For a given model, we assume there is a single fixed $r > 0$. The sequence $\{\gamma_{n,h} : n \geq 1\}$ is defined such that under $\{\gamma_{n,h} : n \geq 1\}$, the asymptotic distribution of $T_n(\theta_0)$ depends on h and only h.

Assumption B. For some $r > 0$, all $h \in H$, all sequences $\{\gamma_{n,h} : n \geq 1\}$, and some distributions J_h, $T_n(\theta_0) \to_d J_h$ under $\{\gamma_{n,h} : n \geq 1\}$.

Assumption B holds in a wide variety of examples of interest, see below and Andrews and Guggenberger (2005, 2009a, 2009b). For a fixed value of h, if $\gamma_{n,h}$

does not depend on n, Assumption B is a standard assumption in the subsampling literature. For example, it is imposed in the basic theorem in PRW, Theorem 2.2.1, p. 43, for subsampling with i.i.d. observations and in Theorem 3.2.1, p. 70, for stationary strong mixing observations. When $\gamma_{n,h}$ does depend on n, the result $T_n(\theta_0) \to_d J_h$ of Assumption B usually can be verified using the same sort of argument as when it does not. In the "continuous limit" case (where Assumption B holds with $p = 0$ and $H = H_2$), the asymptotic distribution J_h may depend on h but is continuous in the sense that one obtains the same asymptotic distribution for any sequence $\{\gamma_{n,h} : n \geq 1\}$ for which $\gamma_{n,h,2}$ converges to $h_2 \in H_2$.

Example 1 (cont.)

In this example, $r = 1/2$ and $H = R_{+,\infty} \times [-1, 1]$ because $\Gamma_1 = R_+$ and $\Gamma_2 = [-1, 1]$. We now verify Assumption B. For more complicated boundary examples, results in Andrews (1999, 2001) can be used to verify Assumption B. The following results are all under the null hypothesis, so the true parameter θ equals zero. For any $h = (h_1, h_2) \in H$ with $h_1 < \infty$ and any sequence $\{\gamma_{n,h} : n \geq 1\}$ of true parameters, consistency of $(\widehat{\sigma}_{n1}, \widehat{\sigma}_{n2}, \widehat{\rho}_n)$ and the CLT imply that

$$\begin{pmatrix} n^{1/2}\overline{X}_{n1}/\widehat{\sigma}_{n1} \\ n^{1/2}\overline{X}_{n2}/\widehat{\sigma}_{n2} \end{pmatrix} \to_d \begin{pmatrix} 0 \\ h_1 \end{pmatrix} + Z_{h_2}, \tag{15}$$

where $Z_{h_2} = (Z_{h_2,1}, Z_{h_2,2})' \sim N(0, V_{h_2})$ and V_{h_2} is a 2×2 matrix with diagonal elements 1 and off-diagonal elements h_2. (For this and the results below, we assume that $\widehat{\sigma}_{n1}$, $\widehat{\sigma}_{n2}$, and $\widehat{\rho}_n$ are consistent in the sense that $\widehat{\sigma}_{nj}/\sigma_{j,n,h} \to_p 1$ for $j = 1, 2$ and $\widehat{\rho}_n - \rho_{n,h} \to_p 0$ under $\{\gamma_{n,h} = (\mu_{n,h}/\sigma_{2,n,h}, \rho_{n,h}, (\sigma_{1,n,h}, \sigma_{2,n,h}, F_{n,h})) : n \geq 1\}$, where $\sigma_{j,n,h}$ denotes σ_j for $j = 1, 2$ and $\rho_{n,h}$ denotes ρ when $\gamma = \gamma_{n,h}$.)

By the continuous mapping theorem, we obtain

$$T_n^*(\theta_0) = n^{1/2}\widehat{\theta}_n/\widehat{\sigma}_{n1} = n^{1/2}\overline{X}_{n1}/\widehat{\sigma}_{n1} - \widehat{\rho}_n \min(0, n^{1/2}\overline{X}_{n2}/\widehat{\sigma}_{n2})) \to_d J_h^* \tag{16}$$

under $\{\gamma_{n,h}\}$, where J_h^* is the distribution of

$$Z_{h_2,1} - h_2 \min(0, Z_{h_2,2} + h_1). \tag{17}$$

Note that J_h^* is stochastically increasing (decreasing) in h_1 for $h_2 < 0$ ($h_2 \geq 0$). Likewise, $-J_h^*$ is stochastically decreasing (increasing) in h_1 for $h_2 < 0$ ($h_2 \geq 0$). (If $Y \sim J_h^*$, then by definition, $-Y \sim -J_h^*$ and $|Y| \sim |J_h^*|$.)

For $h \in H$ with $h_1 = \infty$, we have $\widehat{\theta}_n = \overline{X}_{n1}$ with probability that goes to one (wp$\to$1) under $\{\gamma_{n,h}\}$ because $n^{1/2}\overline{X}_{n2}/\widehat{\sigma}_{n2} \to_p \infty$ under $\{\gamma_{n,h}\}$. (The latter holds because $n^{1/2}\gamma_{n,h,1} = n^{1/2}\mu_{n,h}/\sigma_{2,n,h} \to \infty$, $n^{1/2}(\overline{X}_{n2} - E\overline{X}_{n2})/\widehat{\sigma}_{n2} = O_p(1)$ by the CLT and $\widehat{\sigma}_{n2}/\sigma_{n2} \to_p 1$, and $n^{1/2}E\overline{X}_{n2}/\widehat{\sigma}_{n2} = n^{1/2}\mu_n/\widehat{\sigma}_{n2} \to_p \infty$.) Therefore, under $\{\gamma_{n,h}\}$ with $h_1 = \infty$, we have

$$T_n^*(\theta_0) \to_d J_\infty^*, \quad \text{where } J_\infty^* \text{ is the } N(0,1) \text{ distribution.} \tag{18}$$

Note that J_h^* and J_∞^* do not depend on $\gamma_3 = (\sigma_1, \sigma_2, F)$.

For $T_n(\theta_0) = T_n^*(\theta_0), -T_n^*(\theta_0)$, and $|T_n^*(\theta_0)|$, we have $T_n(\theta_0) \rightarrow_d J_h$ under $\{\gamma_{n,h}\}$, where $J_h = J_h^*, -J_h^*$, and $|J_h^*|$, respectively. Hence, Assumption B holds for upper, lower, and symmetric tests.

3.3. Subsampling Assumptions

For subsampling tests, we require the following additional assumptions.

Assumption C. (i) $b \rightarrow \infty$ and (ii) $b/n \rightarrow 0$.

Assumption D. (i) $\{T_{n,b,j}(\theta_0) : j = 1, \ldots, q_n\}$ are identically distributed under any $\gamma \in \Gamma$ for all $n \geq 1$ and (ii) $T_{n,b,j}(\theta_0)$ and $T_b(\theta_0)$ have the same distribution under any $\gamma \in \Gamma$ for all $n \geq 1$.

Assumption E. For all sequences $\{\gamma_n \in \Gamma : n \geq 1\}$, $U_{n,b}(x) - E_{\theta_0,\gamma_n} U_{n,b}(x) \rightarrow_p 0$ under $\{\gamma_n : n \geq 1\}$ for all $x \in R$.

Assumption F. For all $\varepsilon > 0$ and $h \in H$, $J_h(c_h(1-\alpha)+\varepsilon) > 1-\alpha$, where $c_h(1-\alpha)$ is the $1-\alpha$ quantile of J_h.

Assumption G. For all $h = (h_1, h_2) \in H$ and all sequences $\{\gamma_{n,h} : n \geq 1\}$ for which $b^r \gamma_{n,h,1} \rightarrow g_1$ for some $g_1 \in R_\infty^p$, if $U_{n,b}(x) \rightarrow_p J_g(x)$ under $\{\gamma_{n,h} : n \geq 1\}$ for all $x \in C(J_g)$ for $g = (g_1, h_2) \in R_\infty^{p+q}$, then $L_{n,b}(x) - U_{n,b}(x) \rightarrow_p 0$ under $\{\gamma_{n,h} : n \geq 1\}$ for all $x \in C(J_g)$.

Assumptions C and D are standard assumptions in the subsampling literature (e.g., see PRW, Thm. 2.2.1, p. 43) and are not restrictive. Assumption D necessarily holds when the observations are i.i.d. or stationary and the subsamples are constructed in the usual way. It also holds in some cases with nonstationary observations, such as in the unit root example, see Andrews and Guggenberger (2009a).

Assumption E holds quite generally. For i.i.d.observations, the condition in Assumption E when γ_n does not depend on n is verified in PRW, p. 44, using a U-statistic inequality of Hoeffding. It also holds for any triangular array of row-wise i.i.d. [0,1]-valued random variables by the same argument. Hence, Assumption E holds automatically when the observations are i.i.d. for each fixed $\gamma \in \Gamma$.

For stationary strong mixing observations, the condition in Assumption E when γ_n does not depend on n is verified in PRW, pp. 71–72, by establishing L^2 convergence using a strong mixing covariance bound. It holds for any sequence $\{\gamma_n \in \Gamma : n \geq 1\}$ and, hence, Assumption E holds, by the same argument as in PRW, provided the observations are stationary and strong mixing for each $\gamma \in \Gamma$ and $\sup_{\gamma \in \Gamma} \alpha_\gamma(m) \rightarrow 0$ as $m \rightarrow \infty$, where $\{\alpha_\gamma(m) : m \geq 1\}$ are the strong mixing numbers of the observations when the true parameters are (θ_0, γ).

Assumption F is not restrictive. It holds in all of the examples that we have considered. Assumption G holds automatically when $\{\widehat{T}_{n,b,j}\}$ satisfy Assumption Sub2. In Section 7, we give sufficient conditions for Assumption G when Assumption Sub1 holds.

Example 1 (cont.)

We now verify Assumptions C–F for this example. Assumption G is verified in Section 7 below. We assume b is chosen such that Assumption C holds. Assumption D holds by the i.i.d. assumption. Assumption E holds by the general argument given above for i.i.d. observations. For $\alpha < 1/2$, Assumption F holds for $J_h = J_h^*$ (defined above in (17)–(18)) because for $h_2 \neq -1$, $J_h^*(x)$ is strictly increasing for positive x and $J_h^*(0) = 1/2$. For $h_2 = -1$, $J_h^*(x)$ is strictly increasing for $x \leq h_1$ and $J_h^*(x) = 1$ for $x \geq h_1$. Assumption F holds by analogous reasoning for $J_h = -J_h^*$. Finally, it holds for $J_h = |J_h^*|$ because $|J_h^*(x)|$ is strictly increasing in x for all $h_2 \in [-1, 1]$ (where for $|h_2| = 1$, $|J_h^*(x)|$ has a jump at $x = h_1$ of height $\Pr(Z \geq h_1)$ for $Z \sim N(0,1)$).

4. ASYMPTOTIC RESULTS

Theorem 1 below shows that the asymptotic size of a subsampling test is determined by the asymptotic distributions of the full-sample statistic $T_n(\theta_0)$ and the subsample statistic $T_{n,b,j}(\theta_0)$ under sequences $\{\gamma_{n,h} : n \geq 1\}$. By Assumption B, the asymptotic distribution of $T_n(\theta_0)$ is J_h. The asymptotic distribution of $T_{n,b,j}(\theta_0)$ under $\{\gamma_{n,h} : n \geq 1\}$ is shown to be J_g for some $g \in H$. Given $h \in H$, under $\{\gamma_{n,h} : n \geq 1\}$ not all $g \in H$ are possible indices for the asymptotic distribution of $T_{n,b,j}(\theta_0)$. The set of all possible pairs of localization parameters (g,h) is denoted GH and is defined by

$$\begin{aligned} GH = \{&(g,h) \in H \times H : g = (g_1, g_2),\ h = (h_1, h_2),\ g_2 = h_2, \text{ and for} \\ &m = 1, \ldots, p,\ \text{(i) } g_{1,m} = 0 \text{ if } |h_{1,m}| < \infty,\ \text{(ii) } g_{1,m} \in R_{+,\infty} \text{ if } h_{1,m} \\ &= +\infty, \text{ and (iii) } g_{1,m} \in R_{-,\infty} \text{ if } h_{1,m} = -\infty\}, \end{aligned} \quad \mathbf{(19)}$$

where $g_1 = (g_{1,1}, \ldots, g_{1,p})' \in H_1$ and $h_1 = (h_{1,1}, \ldots, h_{1,p})' \in H_1$. Note that for $(g,h) \in GH$, we have $|g_{1,m}| \leq |h_{1,m}|$ for all $m = 1, \ldots, p$. In the "continuous limit" case (where there is no γ_1 component of γ), GH simplifies considerably: $GH = \{(g_2, h_2) \in H_2 \times H_2 : g_2 = h_2\}$.

The set GH is a crucial ingredient to the asymptotic size of a subsampling test. We now give a simple explanation of its form. Consider the case in which no parameters γ_2, γ_3, and h_2 appear (where $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ and $h = (h_1, h_2)$), $p = 1$, and $\gamma = \gamma_1 \geq 0$. Then, $H = R_{+,\infty}$ and $GH = \{(g,h) \in H \times H : \text{(i) } g = 0 \text{ if } h < \infty \ \&\ \text{(ii) } g \in [0, \infty] \text{ if } h = \infty\}$. The asymptotic distribution of $T_n(\theta_0)$ under the combinations of sample sizes and true values given by $\{(n, \gamma_{n,h}) : n \geq 1\}$ is J_h when $n^r \gamma_{n,h} \to h$ by Assumption B. The question is "What is the asymptotic distribution of $T_{n,b,j}(\theta_0)$ under $\{\gamma_{n,h} : n \geq 1\}$?"

Because $T_{n,b,j}(\theta_0)$ has the same distribution as $T_b(\theta_0)$ for all j by Assumption D, it suffices to determine the asymptotic distribution of $T_b(\theta_0)$ under $\{\gamma_{n,h} : n \geq 1\}$. By a subsequence argument, the asymptotic distribution of $T_b(\theta_0)$ under the combinations of sample sizes and true values given by $\{(b, \gamma_{n,h}) : n \geq 1\}$ is J_g when $b^r \gamma_{n,h} \to g$ by Assumption B. If $n^r \gamma_{n,h} \to h < \infty$, then $b^r \gamma_{n,h} \to g = 0$ because $b/n \to 0$ by Assumption C. Hence, if $h < \infty$, then $g = 0$, which is the first condition in GH (defined in the previous paragraph). On the other hand, for any $g \in [0, \infty]$, there exists a sequence $\{\gamma_{n,h} : n \geq 1\}$ such that $n^r \gamma_{n,h} \to \infty$ and $b^r \gamma_{n,h} \to g$. This explains the second condition in GH.

We return now to the general case. Consistent with the heuristics above, Theorem 1 below shows that for a subsampling test $AsySz(\theta_0) \in [Max_{Sub}(\alpha), Max^-_{Sub}(\alpha)]$, where

$$Max_{Sub}(\alpha) = \sup_{(g,h)\in GH} [1 - J_h(c_g(1-\alpha))] \quad \text{and}$$
$$Max^-_{Sub}(\alpha) = \sup_{(g,h)\in GH} [1 - J_h(c_g(1-\alpha)-)]. \tag{20}$$

Define $Min_{Sub}(\alpha)$ and $Min^-_{Sub}(\alpha)$ analogously with “inf” in place of “sup.” In the “continuous limit” case, $Max_{Sub}(\alpha)$ simplifies to $\sup_{h\in H}[1 - J_h(c_h(1-\alpha))]$, which is less than or equal to α by the definition of $c_h(1-\alpha)$.

Analogously, for FCV tests, define

$$Max_{Fix}(\alpha) = \sup_{h\in H}[1 - J_h(c_{Fix}(1-\alpha))] \quad \text{and}$$
$$Max^-_{Fix}(\alpha) = \sup_{h\in H}[1 - J_h(c_{Fix}(1-\alpha)-)]. \tag{21}$$

Define $Min_{Fix}(\alpha)$ and $Min^-_{Fix}(\alpha)$ analogously with “inf” in place of “sup.”

THEOREM 1.

(i) Suppose Assumptions A and B hold. Then, an FCV test satisfies

$$AsySz(\theta_0) \in [Max_{Fix}(\alpha), Max^-_{Fix}(\alpha)] \quad \textit{and}$$
$$AsyMinRP(\theta_0) \in [Min_{Fix}(\alpha), Min^-_{Fix}(\alpha)].$$

(ii) Suppose Assumptions A–G hold. Then, a subsampling test satisfies

$$AsySz(\theta_0) \in [Max_{Sub}(\alpha), Max^-_{Sub}(\alpha)] \quad \textit{and}$$
$$AsyMinRP(\theta_0) \in [Min_{Sub}(\alpha), Min^-_{Sub}(\alpha)].$$

Comments

1. If $J_h(x)$ is continuous at the appropriate value(s) of x (which holds in almost all examples), then $Max_{Fix}(\alpha) = Max^-_{Fix}(\alpha)$ and $Max_{Sub}(\alpha) = Max^-_{Sub}(\alpha)$.

In this case, Theorem 1 gives the precise value of $AsySz(\theta_0)$ and analogously for $AsyMinRP(\theta_0)$.

2. A key question concerning nominal level α FCV and subsampling tests is whether $AsySz(\theta_0) \leq \alpha$. For an FCV test with $Max_{Fix}(\alpha) = Max^{-}_{Fix}(\alpha)$, Theorem 1(i) shows that this holds if and only if (iff) $c_{Fix}(1-\alpha)$ is greater than or equal to the $1-\alpha$ quantile of J_h, $c_h(1-\alpha)$, for all $h \in H$.
3. For a subsampling test with $Max^{-}_{Sub}(\alpha) = Max_{Sub}(\alpha)$, Theorem 1(ii) shows that $AsySz(\theta_0) \leq \alpha$ iff $c_g(1-\alpha) \geq c_h(1-\alpha)$ for all $(g,h) \in GH$. In consequence, a graph of $c_h(1-\alpha)$ as a function of h is very informative about the asymptotic size of a subsampling test. Figure 1 provides four examples of shapes of $c_h(1-\alpha)$ as a function of h in the special case where $h = h_1 \in R_{+,\infty}$. In Figure 1(a), $c_h(1-\alpha)$ is decreasing in h. Hence, $c_g(1-\alpha) \geq c_h(1-\alpha)$ for all $(g,h) \in GH$ (since $g \leq h$) and $AsySz(\theta_0) \leq \alpha$. A decreasing quantile graph occurs in (i) the nuisance parameter near a boundary example for one-sided tests for one sign of a correlation parameter (but not the other sign), see Section 6.1; (ii) the weak instruments regression example for tests concerning the coefficient on an endogenous variable for one-sided tests for one sign of a correlation and for symmetric two-sided tests for all values of a correlation not close to zero, see Andrews and Guggenberger (2005); (iii) the moment inequality example, see Andrews and Guggenberger (2009b), and (iv) the autoregressive example for lower one-sided and symmetric two-sided tests, see Andrews and Guggenberger (2009a). (In all of these examples except the autoregressive example, a parameter h_2 appears, which corresponds to a correlation parameter, and the graphs described are actually for h_1 for a given value of h_2.)

   In contrast, in Figures 1(b), 1(c), and 1(d), there are pairs $(g,h) \in GH$ for which $c_g(1-\alpha) < c_h(1-\alpha)$ for $g < h$ and, hence, $AsySz(\theta_0) > \alpha$. Figure 1(c) illustrates a case in which subsampling does not lead to over-rejection for alternatives that typically are contiguous to $h = 0$, i.e., those with $h < \infty$, but leads to over-rejection for alternatives that typically are not contiguous, i.e., $h = \infty$.

   An increasing quantile graph, as in Figure 1(b), occurs in (i) the nuisance parameter near a boundary example for one-sided tests for the other sign of the correlation parameter than in the case described above; (ii) the weak instruments regression example referred to above for one-sided tests for the other sign of the correlation; and (iii) the autoregressive example for upper one-sided tests. The bowl shape of Figure 1(c) occurs in the weak instruments regression example referred to above for symmetric two-sided tests for values of the correlation very close to zero. The hump shape of Figure 1(d) occurs in the post-conservative model selection example with the height of the hump depending on a correlation parameter, see Andrews and Guggenberger (2009a).

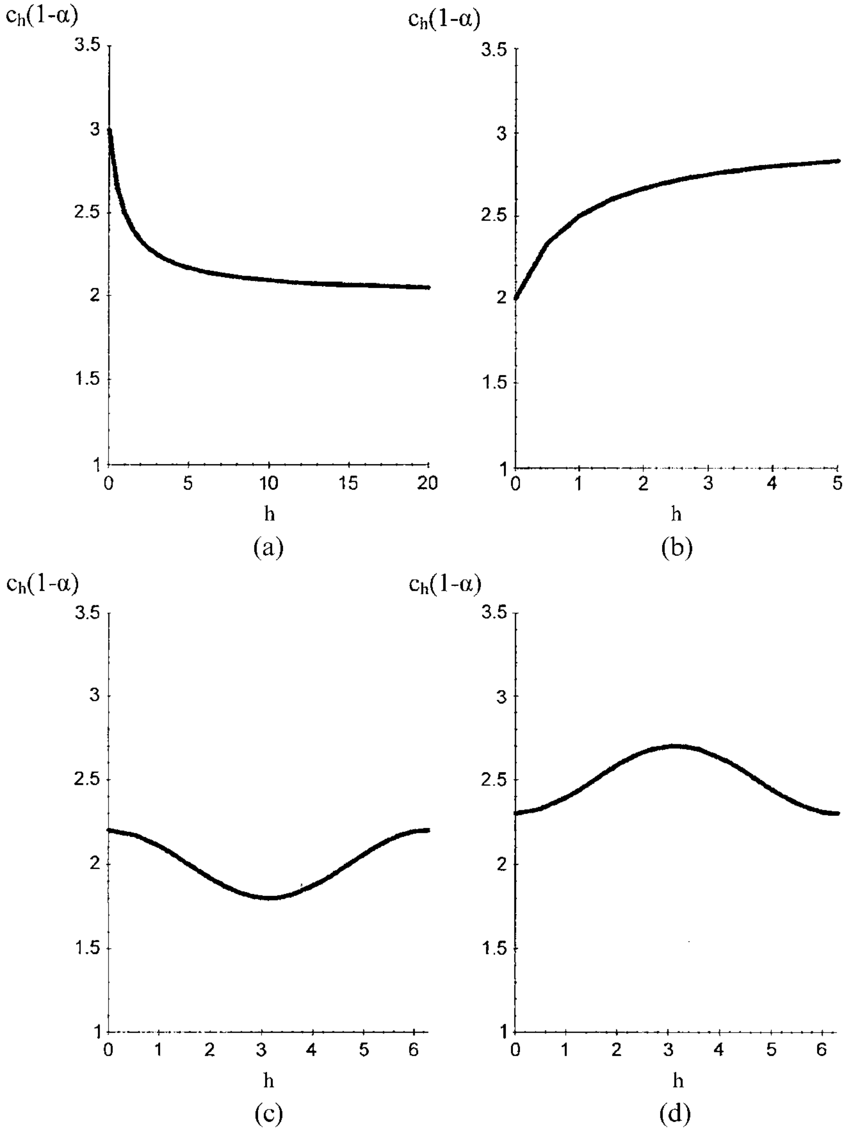

FIGURE 1. $1-\alpha$ quantile of J_h, $c_h(1-\alpha)$, as a function of h.

4. The same argument as used to prove Theorem 1 shows that, for FCV and subsampling tests, $ExSz_n(\theta_0) \to [Max_{Type}(\alpha), Max^-_{Type}(\alpha)]$ for $Type = Fix$ and Sub, respectively. Hence, when $Max_{Type}(\alpha) = Max^-_{Type}(\alpha)$, we have $\lim_{n\to\infty} ExSz_n(\theta_0) = Max_{Type}(\alpha)$ for $Type = Fix$ and Sub.

5. Andrews and Guggenberger (2009a) utilizes the results of Theorem 1 to introduce and analyze various new procedures including (i) hybrid subsampling/FCV, (ii) size-corrected FCV, (iii) size-corrected subsampling, and (iv) size-corrected hybrid tests and CIs (and analogous m out of n bootstrap procedures). These procedures extend the applicability of subsampling, m out of n bootstrap, and FCV methods to a wide variety of models whose asymptotic distributions are discontinuous in some parameter.

5. EQUAL-TAILED *t*-TESTS

This section considers *equal-tailed* two-sided t-tests. It is of interest to see how the asymptotic size properties of equal-tailed tests compare to those of symmetric tests. In short, it turns out that in many examples, equal-tailed subsampling tests perform worse.

Suppose $T_n(\theta_0)$ is a t statistic. A nominal level $\alpha(\in(0,1/2))$ equal-tailed t-test of $H_0:\theta=\theta_0$ versus $H_1:\theta\neq\theta_0$ rejects H_0 when

$$T_n(\theta_0) > c_{1-\alpha/2} \quad \text{or} \quad T_n(\theta_0) < c_{\alpha/2}, \tag{22}$$

where $c_{1-\alpha}=c_{Fix}(1-\alpha)$ for FCV tests and $c_{1-\alpha}=c_{n,b}(1-\alpha)$ for subsampling tests.

The exact size, $ExSz_n(\theta_0)$, of the equal-tailed t-test is

$$ExSz_n(\theta_0) = \sup_{\gamma\in\Gamma}\left(P_{\theta_0,\gamma}(T_n(\theta_0) > c_{1-\alpha/2}) + P_{\theta_0,\gamma}(T_n(\theta_0) < c_{\alpha/2})\right). \tag{23}$$

The asymptotic size of the test is $AsySz(\theta_0)=\limsup_{n\to\infty}ExSz_n(\theta_0)$. For brevity, we only state results for the $AsySz(\theta_0)$ of subsampling tests. Results for $AsyMinRP(\theta_0)$ and FCV tests are analogous. Define

$$Max^{r-}_{ET,Sub}(\alpha) = \sup_{(g,h)\in GH}[1-J_h(c_g(1-\alpha/2))+J_h(c_g(\alpha/2)-)] \quad \text{and}$$

$$Max^{\ell-}_{ET,Sub}(\alpha) = \sup_{(g,h)\in GH}[1-J_h(c_g(1-\alpha/2)-)+J_h(c_g(\alpha/2))]. \tag{24}$$

Here "$r-$" denotes that the limit from the left "$-$" appears in the right summand.

Assumption F is replaced by Assumption J.

Assumption J. For all $\varepsilon>0$ and $h\in H$, $J_h(c_h(\tau)+\varepsilon)>\tau$ for $\tau=\alpha/2$ and $\tau=1-\alpha/2$, where $c_h(\tau)$ is the τ quantile of J_h.

The proof of Theorem 1 can be adjusted straightforwardly to yield the following result.

COROLLARY 2. *Let $\alpha\in(0,1/2)$ be given. Suppose Assumptions A–E, G, and J hold. Then, an equal-tailed subsampling test satisfies*

$$AsySz(\theta_0)\in[Max^{r-}_{ET,Sub}(\alpha), Max^{\ell-}_{ET,Sub}(\alpha)].$$

Comments

1. If $J_h(x)$ is continuous at the appropriate value(s) of x, then $Max^{r-}_{ET,Sub}(\alpha) = Max^{\ell-}_{ET,Sub}(\alpha) = AsySz(\theta_0)$.
2. By Corollary 2 and the definition of $Max^{\ell-}_{ET,Sub}(\alpha)$, sufficient conditions for a nominal level α equal-tailed subsampling test to have asymptotic level α are: (i) $c_g(1-\alpha/2) \geq c_h(1-\alpha/2)$ for all $(g,h) \in GH$, (ii) $c_g(\alpha/2) \leq c_h(\alpha/2)$ for all $(g,h) \in GH$, and (iii) $\sup_{h \in H}[1 - J_h(c_h(1-\alpha/2)-) + J_h(c_h(\alpha/2))] = \sup_{h \in H}[1 - J_h(c_h(1-\alpha/2)) + J_h(c_h(\alpha/2)-)]$. Conditions (i) and (ii) automatically hold in "continuous limit" cases. They also hold in some "discontinuous limit" cases, but often fail in such cases. Condition (iii) holds in most examples.

Example 1 (cont.)

In this example, the critical values $(c_{\alpha/2}, c_{1-\alpha/2})$ for the equal-tailed FCV and subsampling tests are $(z_{\alpha/2}, z_{1-\alpha/2})$ and $(c_{n,b}(\alpha/2), c_{n,b}(1-\alpha/2))$, respectively. Next, we verify Assumption J for this example. For $|h_2| < 1$, $J_h(x) = J_h^*(x)$ is strictly increasing for all $x \in R$. When $h_2 = 1$, $J_h(x) = J_h^*(x)$ equals zero for $x < -h_1$ and is strictly increasing for all $x \geq -h_1$. Finally, for $h_2 = -1$, $J_h(x) = J_h^*(x)$ is strictly increasing for all $x \leq h_1$ and equals 1 otherwise. In consequence, Assumption J holds.

6. EXAMPLES

6.1. Test When a Nuisance Parameter May Be Near a Boundary (cont.)

For this example, some calculations (given in Appendix A) show that for upper and lower one-sided and symmetric two-sided tests, $Max^{-}_{Type}(\alpha) = Max_{Type}(\alpha)$ for $Type = Fix$ and Sub. Hence, by Theorem 1, $AsySz(\theta_0) = Max_{Type}(\alpha)$ for these tests. Analogously, for equal-tailed tests, calculations (given in Appendix A) show that $Max^{r-}_{ET,Type}(\alpha) = Max^{\ell-}_{ET,Type}(\alpha)$ for any $\alpha \in (0, 1/2)$ (and the limit from the left "$-$" in both can be deleted) for $Type = Fix$ and Sub. Hence, by Corollary 2, $AsySz(\theta_0) = Max^{r-}_{ET,Type}(\alpha)$ for equal-tailed tests.

Given that $J_h = J_h^*$ is stochastically increasing (decreasing) in h_1 for fixed $h_2 < 0$ ($h_2 \geq 0$), some calculations (given in Appendix A) yield the following simplifications for upper one-sided tests:

$$Max_{Fix}(\alpha) = \sup_{h \in H}[1 - J_h(c_{Fix}(1-\alpha))] = \sup_{h_2 \in [0,1]} (1 - J_{(0,h_2)}(z_{1-\alpha})) \quad \text{and}$$

$$Max_{Sub}(\alpha) = \sup_{(g,h) \in GH}[1 - J_h(c_g(1-\alpha))] = \sup_{h_2 \in [-1,0]} (1 - J_\infty(c_{(0,h_2)}(1-\alpha))), \tag{25}$$

where $J_{(0,h_2)}$ is the distribution of $Z_{h_2,1} - h_2 \min(0, Z_{h_2,2})$, $(Z_{h_2,1}, Z_{h_2,2})$ is bivariate normal with means zero, variances one, and correlation h_2, and J_∞ is the standard normal distribution. The results for lower one-sided tests are analogous with $h_2 \in [0,1]$ and $h_2 \in [-1,0]$ replaced by $h_2 \in [-1,0]$ and $h_2 \in [0,1]$, respectively.

Figure 2 provides .95 quantile graphs of J_h^* and $|J_h^*|$ as functions of $h_1 \geq 0$ for several values of $h_2 \in [-1,1]$. As discussed in Comment 3 to Theorem 1, these graphs provide considerable qualitative information concerning the null rejection probabilities of subsampling and FCV tests as a function of h_1 $(= \lim_{n\to\infty} n^{1/2}\mu_{n,h}/\sigma_{2,n,h})$ and h_2 $(= \lim_{n\to\infty} \rho_{n,h})$. For example, the quantile graphs for J_h^* indicate that the upper one-sided subsampling test over-rejects for negative values of h_2 for all (g_1, h_1) pairs with $g_1 < h_1$ (because the graphs are increasing in h_1), with the greatest degree of over-rejection being quite large and occurring for $(g_1, h_1) = (0, \infty)$ and h_2 close to -1. On the other hand, for positive values of h_2, the upper subsampling test underrejects (because the graph is decreasing in h_1), with the greatest degree of underrejection being relatively small and occurring for $(g_1, h_1) = (0, \infty)$ and h_2 around .5. In sum, the quantile graphs indicate qualitatively that the size of the upper subsampling test exceeds .05 by a substantial amount.

Table 1 provides quantitative information concerning the size properties of the tests in this example. It is obtained by calculating asymptotic quantities by simulation. Table 1 reports $AsySz(\theta_0)$ as well as the maximum asymptotic null rejection probabilities ($\times 100$) for given h_2 for a range of h_2 values in $[-1,1]$ and $\alpha = .05$.

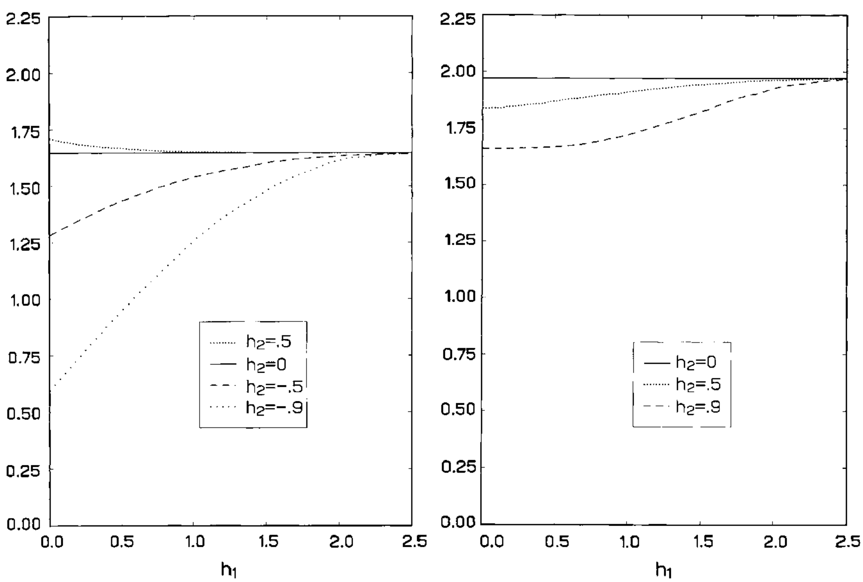

FIGURE 2. Nuisance parameter near a boundary example: .95 quantile graphs, $c_h(.95)$, for J_h^* and $\left|J_h^*\right|$ as functions of h_1 for several values of h_2.

TABLE 1. Nuisance parameter near a boundary example: Maximum asymptotic null rejecton probabilities ($\times 100$) as a function of the true correlation h_2 for nominal 5% tests

	Upper 1-sided		Symmetric 2-sided		Equal-tailed 2-sided	
h_2	Sub	FCV	Sub	FCV	Sub	FCV
−1.00	50.2	5.0	10.0	5.0	52.5	5.0
−.99	42.8	5.0	10.0	5.0	43.1	5.0
−.95	33.8	5.0	10.0	5.0	32.4	5.0
−.90	27.6	5.0	10.0	5.0	25.3	5.0
−.80	20.2	5.0	9.5	5.0	17.4	5.0
−.60	12.3	5.0	7.5	5.0	10.0	5.0
−.40	8.3	5.0	6.0	5.0	6.8	5.0
−.20	6.2	5.0	5.2	5.0	5.4	5.0
.00	5.0	5.0	5.0	5.0	5.0	5.0
.20	5.0	5.6	5.2	5.0	5.4	5.0
.40	5.0	5.8	6.0	5.0	6.8	5.0
.60	5.0	5.6	7.5	5.0	10.0	5.0
.80	5.0	5.1	9.5	5.0	17.4	5.0
.90	5.0	5.0	10.0	5.0	25.3	5.0
.95	5.0	5.0	10.0	5.0	32.4	5.0
.99	5.0	5.0	10.0	5.0	43.1	5.0
1.00	5.0	5.0	10.0	5.0	52.5	5.0
$AsySz(\theta_0)$	50.2	5.8	10.0	5.0	52.5	5.0

(These maxima are over $h \in H$ for FCV tests and $(g,h) \in GH$ for subsampling tests with h_2 fixed. For example, for upper one-sided FCV and subsampling tests, these maxima simplify to $1 - J_{(0,h_2)}(z_{1-\alpha})$ and $1 - J_\infty(c_{(0,h_2)}(1-\alpha))$, respectively. For the two-sided FCV test, the maximum is $\sup_{h_1 \in [0,\infty]}[1 - J_{(h_1,h_2)}(z_{1-\alpha/2})]$.) Results for lower one-sided tests are not reported in Table 1 because they are the same as those for upper one-sided tests, but with h_2 replaced by $-h_2$. The simulations use 50,000 repetitions, and when maximization over h_1 is needed the upper bound is 12 and a grid of size 0.05 is used. The last row of Table 1 gives the $AsySz(\theta_0)$ of each test, which is maximum of the numbers in each column.

The results of Table 1 are summarized as follows: For upper one-sided tests, we find large asymptotic size distortions for the subsampling tests and very small size distortions for the FCV tests for all nominal sizes $\alpha \in [.01, .2]$ that we consider. (Only results for $\alpha = .05$ are reported.) The upper one-sided subsampling test overrejects the null when the correlation h_2 is negative, does not overreject when h_2 is positive, and the magnitude of overrejection increases as h_2 gets closer to -1. This test has asymptotic size approximately equal to $1/2$ for all nominal sizes $\alpha \in [.01, .2]$ that we consider.

The symmetric two-sided subsampling test also is found to be size-distorted asymptotically, but by a much smaller amount. The Monte Carlo simulations for $\alpha \in [.01, .2]$ show that $AsySz(\theta_0)$ is approximately 2α for the symmetric subsampling test. Its rejection rate is invariant to the sign of h_2. In contrast, the two-sided FCV test is found to have asymptotic size equal to its nominal level, although this test is not asymptotically similar in a uniform sense.

The equal-tailed subsampling test is found to have a large asymptotic size distortion: $AsySz(\theta_0)$ is approximately $1/2+\alpha/2 = .525$. The two-sided FCV test has no asymptotic size distortion, but is not asymptotically similar in a uniform sense.

6.2. Tests Concerning an Exogenous Variable in an IV Regression Model with Possibly Weak Instruments

We consider the following IV regression model:

$$y_1 = y_2\beta + X\theta + u,$$
$$y_2 = \widetilde{Z}\pi + X\widetilde{\phi} + v, \tag{26}$$

where $y_1, y_2 \in R^n$ are endogenous variable vectors, $X \in R^n$ is an exogenous variable vector, $\widetilde{Z} \in R^{n\times k_2}$ for $k_2 \geq 1$ is a matrix of IVs, and $(\beta, \theta, \widetilde{\phi}, \pi')' \in R^{1\times 1\times 1\times k_2}$ are unknown parameters. Denote by u_i, v_i, X_i, and $\widetilde{Z}_i$ the ith rows of u, v, X, and $\widetilde{Z}$, respectively, written as column vectors (or scalars). Let $k = 1+k_2$.

We are interested in tests concerning the parameter θ on the exogenous variable X in the equation for y_1 when the IVs $\widetilde{Z}$ may be weak (i.e., π may be close to 0). The null hypothesis is

$$H_0 : \theta = \theta_0. \tag{27}$$

The alternative hypothesis may be one-sided or two-sided. Below we consider FCV and subsampling tests based on the two-stage least squares (2SLS) estimator. We consider upper and lower one-sided and symmetric and equal-tailed two-sided tests of nominal level α.

The literature on IV regression models when the IVs may be weak is now voluminous. For surveys, see Stock, Wright, and Yogo (2002), Dufour (2003), Hahn and Hausman (2003), and Andrews and Stock (2007). Most of the papers in the literature focus on tests that concern the coefficient β on the endogenous variable y_2. For example, the subsampling results in Andrews and Guggenberger (2005) consider this null and differ from the results given here in this respect. Kleibergen (2008) is one paper that focuses on tests of $H_0 : \theta = \theta_0$. It considers different test statistics from the 2SLS-based statistics that are considered here. Given the focus here on the 2SLS estimator, the results below are closely related to results in Staiger and Stock (1997). In fact, part of the proofs utilizes results from Staiger and Stock (1997).

Define

$$Z_i = \widetilde{Z}_i - (\mathrm{E}\widetilde{Z}_i X_i)(\mathrm{E}X_i^2)^{-1}X_i \quad \text{and} \quad \phi = \widetilde{\phi} + (\mathrm{E}X_i^2)^{-1}(\mathrm{E}X_i\widetilde{Z}_i')\pi. \tag{28}$$

The reduced-form equation for y_2 can be rewritten as

$$y_{2i} = Z_i'\pi + X_i\phi + v_i. \tag{29}$$

By construction, $EZ_iX_i = 0$. Note that Z_i is unobserved. As explained below, for our purposes, this does not matter.

Let $\overline{Z} = [X{:}Z]$, where $Z = [Z_1 \cdots Z_n]'$. For any matrix M we let $P_M = M(M'M)^{-1}M'$, and for any conformable matrix M we let $M^{\perp} = M - P_X M$.

We define a partially-studentized test statistic $T_n^*(\theta_0)$ based on the 2SLS estimator, $\widehat{\theta}_n$, of θ as follows:

$$T_n^*(\theta_0) = \frac{\widehat{\theta}_n - \theta_0}{\widehat{\sigma}_n}, \quad \text{where} \quad \widehat{\theta}_n = (X'X)^{-1}X'(y_1 - y_2\widehat{\beta}_n), \quad \widehat{\beta}_n = \frac{y_2' P_{Z^{\perp}} y_1}{y_2' P_{Z^{\perp}} y_2},$$
$$\widehat{\sigma}_n = (X'SX)^{-1/2}, \quad \text{and} \quad S = P_{\overline{Z}} - \frac{P_{\overline{Z}} y_2 y_2' P_{\overline{Z}}}{y_2' P_{\overline{Z}} y_2}. \tag{30}$$

Note that $T_n^*(\theta_0)$ is unchanged if Z is replaced by $\widetilde{Z}$ in its definition because $P_{\overline{Z}} = P_{[X:\widetilde{Z}]}$ and $P_{Z^{\perp}} = P_{\widetilde{Z}^{\perp}}$. The statistic $T_n^*(\theta_0)$ is only partially studentized because it does not employ an estimator of $\sigma_u = Eu_i^2$. The standard fully-studentized test statistic is

$$T_n^*(\theta_0)/\widehat{\sigma}_u, \quad \text{where} \quad \widehat{\sigma}_u^2 = (n-1)^{-1}(y_1^{\perp} - y_2^{\perp}\widehat{\beta}_n)'(y_1^{\perp} - y_2^{\perp}\widehat{\beta}_n). \tag{31}$$

Standard nominal-level α 2SLS tests based on a fixed critical value (FCV) employ the test statistic $T_n(\theta_0)/\widehat{\sigma}_u$, where $T_n(\theta_0) = T_n^*(\theta_0), -T_n^*(\theta_0)$, and $|T_n^*(\theta_0)|$ for upper one-sided, lower one-sided, and symmetric two-sided tests, respectively. In each case, the test rejects H_0 if

$$T_n(\theta_0)/\widehat{\sigma}_u > c_\infty(1-\alpha), \tag{32}$$

where $c_\infty(1-\alpha) = z_{1-\alpha}$, $z_{1-\alpha}$, and $z_{1-\alpha/2}$, respectively, and $z_{1-\alpha}$ denotes the $1-\alpha$ quantile of the standard normal distribution. For FCV tests, full studentization of the test statistic is necessary for the normal critical values to be suitable when the IVs are strong.

Next, we consider subsampling tests. Subsampling tests can be based on the partially-studentized statistic $T_n^*(\theta_0)$ or the fully-studentized statistic $T_n^*(\theta_0)/\widehat{\sigma}_u$. We focus on $T_n^*(\theta_0)$ but provide some results for $T_n^*(\theta_0)/\widehat{\sigma}_u$ as well. The rationale for using $T_n^*(\theta_0)$ is that σ_u^2 is difficult to estimate when the IVs are weak and a subsampling test does not require normalization for the scale of the error because the subsample statistics have the same error scale as the full-sample statistic.

Let $\{T_{n,b,j}^*(\theta_0) : j = 1, \ldots, q_n\}$ be partially-studentized subsample t statistics that are defined just as $T_n^*(\theta_0)$ is defined but are based on the jth subsample of length b. That is, $T_{n,b,j}^*(\theta_0) = (\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}$, where $\widehat{\theta}_{n,b,j}$ and $\widehat{\sigma}_{n,b,j}$ are analogues of $\widehat{\theta}_n$ and $\widehat{\sigma}_n$, respectively, based on the jth subsample. Note that the subsample statistic $T^*_{n,b,j}(\theta_0)$ is centered at θ_0, rather than $\widehat{\theta}_n$, and hence Assumption Sub2 holds. This choice of centering is made because $\widehat{\theta}_n$ is not consistent if the IVs are weak and ϕ is local to zero, hence, centering at $\widehat{\theta}_n$ would yield poor performance of the subsampling test.

The nominal level α subsampling test rejects $H_0 : \theta = \theta_0$ if

$$T_n(\theta_0) > c_{n,b}(1-\alpha), \tag{33}$$

where $T_n(\theta_0) = T^*_n(\theta_0), -T^*_n(\theta_0)$, and $|T^*_n(\theta_0)|$ and $c_{n,b}(1-\alpha)$ denotes the subsampling critical value defined in (6). Equal-tailed two-sided subsampling tests are defined in (22) with $c_{1-\alpha} = c_{n,b}(1-\alpha)$.

Neither the distribution of $\widehat{\theta}_n - \theta_0$ nor that of $\widehat{\sigma}_n$ depend on θ_0 when θ_0 is the true value. Therefore, the finite-sample distribution of $T^*_n(\theta_0)$ under $H_0 : \theta = \theta_0$ does not depend on θ_0 and the test results given below for fixed θ_0 hold uniformly over $\theta_0 \in R$. This implies that the test results apply immediately to CIs constructed by inverting the tests.

6.2.1. Assumptions and Parameter Space. We assume that $\{(u_i, v_i, X_i, Z_i) : i \le n\}$ are i.i.d. with distribution F. We define a vector of nuisance parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ as follows: Let $\gamma_1 = (\gamma_{11}, \gamma_{12})' \in R^2$. Define

$$\gamma_{11} = ||Q^{1/2}_{ZZ}\pi/\sigma_v||, \quad \gamma_{12} = Q^{1/2}_{XX}\phi/\sigma_v, \quad \gamma_2 = \rho = Corr_F(u_i, v_i), \quad \text{and}$$

$$\gamma_3 = (F, \pi, \phi), \quad \text{where} \quad \sigma^2_v = \mathrm{E}_F v^2_i,\ \sigma^2_u = \mathrm{E}_F u^2_i, \quad \text{and}$$

$$Q = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix} = \mathrm{E}_F \overline{Z}_i \overline{Z}'_i. \tag{34}$$

We choose this specification for γ_{11}, γ_{12}, and γ_2 because the asymptotic distribution of the t statistic depends only on these scalar parameters, as shown below.

The parameter space for γ_1 is $\Gamma_1 = R_+ \times R$. We specify the parameter space for γ_2 to be $\Gamma_2 = \Gamma_{2,\rho_U} = [-\rho_U, \rho_U]$ for some $\rho_U \in (0, 1]$. We allow for different bounds ρ_U because we are interested in how the asymptotic sizes of FCV and subsampling tests vary with the upper bound ρ_U. For given $(\gamma_1, \gamma_2) \in \Gamma_1 \times \Gamma_{2,\rho_U}$, the parameter space for γ_3 is

$$\begin{aligned}
&\Gamma_3(\gamma_1, \gamma_2) \\
&\quad = \left\{ (F, \pi, \phi) : \mathrm{E}_F u^2_i = \sigma^2_u, \mathrm{E}_F v^2_i = \sigma^2_v, \mathrm{E}_F \overline{Z}_i \overline{Z}'_i = Q = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix}, \right. \\
&\qquad \& \ \mathrm{E}_F u_i v_i/(\sigma_u \sigma_v) = \rho \text{ for some } \sigma^2_u, \sigma^2_v > 0,\ \text{pd } Q \in R^{k \times k},\ \rho \in [-1, 1], \\
&\qquad \pi \in R^{k_2},\ \& \ \phi \in R \text{ that satisfy } \left\| Q^{1/2}_{ZZ}\pi/\sigma_v \right\| = \gamma_{11},\ Q^{1/2}_{XX}\phi/\sigma_v = \gamma_{12} \ \&
\end{aligned}$$

$$\rho = \gamma_2; \mathrm{E}_F u_i \overline{Z}_i = \mathrm{E}_F v_i \overline{Z}_i = 0;\ \mathrm{E}_F(u_i^2, v_i^2, u_i v_i)\overline{Z}_i \overline{Z}_i' = (\sigma_u^2, \sigma_v^2, \sigma_u \sigma_v \rho) Q;$$

$$\lambda_{\min}(\mathrm{E}_F \overline{Z}_i \overline{Z}_i') \geq \varepsilon;\ \left\| \mathrm{E}_F \left(|u_i/\sigma_u|^{2+\delta},\ |v_i/\sigma_v|^{2+\delta},\ |u_i v_i/(\sigma_u \sigma_v)|^{2+\delta} \right)' \right\| \leq M,$$

$$\left. \&\ \left\| \mathrm{E}_F \left(||\overline{Z}_i u_i/\sigma_u||^{2+\delta},\ ||\overline{Z}_i v_i/\sigma_v||^{2+\delta},\ ||\overline{Z}_i||^{2+\delta} \right)' \right\| \leq M \right\} \quad \textbf{(35)}$$

for some constants $\varepsilon > 0$, $\delta > 0$, and $M < \infty$, where pd denotes "positive definite." Assumption A holds in this example.

The tests introduced above are equivalent to analogous tests defined with $T_n^*(\theta_0)$, $T_{n,b,j}^*(\theta_0)$, and $\widehat{\sigma}_u$ replaced by

$$T_n^{**}(\theta_0) = T_n^*(\theta_0)/\sigma_u, \quad T_{n,b,j}^{**}(\theta_0) = T_{n,j}^*(\theta_0)/\sigma_u, \quad \text{and } \widehat{\sigma}_u/\sigma_u, \quad \textbf{(36)}$$

respectively. (They are "equivalent" in the sense that they generate the same critical regions. The reason is that for all of the tests above $1/\sigma_u$ scales both the test statistic and the critical value equally, e.g., $T_n^*(\theta_0) > \widehat{\sigma}_u c_\infty(1-\alpha)$ iff $T_n^{**}(\theta_0) > (\widehat{\sigma}_u/\sigma_u) c_\infty(1-\alpha)$.) We determine the $AsySz(\theta_0)$ of the tests written as in (36) because this eliminates σ_u from the asymptotic distributions that arise and, hence, simplifies the expressions.

6.2.2. Asymptotic Distributions. Next, we verify Assumption B for the test statistic $T_n^{**}(\theta_0)$. In this example, $r = 1/2$ and the parameter space H is

$$H = H_{\rho_U} = R_{+,\infty} \times R_\infty \times [-\rho_U, \rho_U]. \quad \textbf{(37)}$$

For $h \in H$, let $\{\gamma_{n,h} : n \geq 1\}$ denote a sequence of parameters in Γ with subvectors $\gamma_{n,h,j}$ for $j = 1, 2, 3$ defined by

$$\begin{aligned} &\gamma_{n,h,1} = (||(\mathrm{E}_{F_n} Z_i Z_i')^{1/2} \pi_n / (\mathrm{E}_{F_n} v_i^2)^{1/2}||, (\mathrm{E}_{F_n} X_i X_i')^{1/2} \phi_n / (\mathrm{E}_{F_n} v_i^2)^{1/2})', \\ &\gamma_{n,h,2} = Corr_{F_n}(u_i, v_i),\ n^{1/2}\gamma_{n,h,1} \to h_1,\ \gamma_{n,h,2} \to h_2, \quad \text{and} \\ &\gamma_{n,h,3} = (F_n, \pi_n, \phi_n) \in \Gamma_3(\gamma_{n,h,1}, \gamma_{n,h,2}). \end{aligned} \quad \textbf{(38)}$$

By the central limit theorem and $\mathrm{E} Z_i X_i = 0$, it follows that

$$\begin{pmatrix} (X'X)^{-1/2} X'u/\sigma_u \\ (X'X)^{-1/2} X'v/\sigma_v \\ (Z'Z)^{-1/2} Z'u/\sigma_u \\ (Z'Z)^{-1/2} Z'v/\sigma_v \end{pmatrix} \to_d \begin{pmatrix} \psi_{Xu,h_2} \\ \psi_{Xv,h_2} \\ \psi_{Zu,h_2} \\ \psi_{Zv,h_2} \end{pmatrix} \sim N\left(0, \begin{pmatrix} V_{h_2} & 0 \\ 0 & V_{h_2} \otimes I_{k_2} \end{pmatrix}\right) \quad \text{for}$$

$$V_{h_2} = \begin{bmatrix} 1 & h_2 \\ h_2 & 1 \end{bmatrix}, \quad \textbf{(39)}$$

where $\psi_{Xu,h_2}, \psi_{Xv,h_2} \in R$, $\psi_{Zu,h_2}, \psi_{Zv,h_2} \in R^{k_2}$, and $h_2 \in [-\rho_U, \rho_U]$.

If $||h_1|| < \infty$, then the IVs are weak, see (38). From (39) and calculations in Andrews and Guggenberger (2005), it follows that under $\{\gamma_{n,h}\}$, we have

$$\begin{pmatrix} y_2' P_{Z^\perp} u/(\sigma_u \sigma_v) \\ y_2' P_{Z^\perp} y_2/\sigma_v^2 \end{pmatrix} \rightarrow_d \begin{pmatrix} \xi_{1,h} \\ \xi_{2,h} \end{pmatrix} = \begin{pmatrix} (\psi_{Zv,h_2} + h_{11} s_{k_2})' \psi_{Zu,h_2} \\ (\psi_{Zv,h_2} + h_{11} s_{k_2})'(\psi_{Zv,h_2} + h_{11} s_{k_2}) \end{pmatrix}, \tag{40}$$

where s_{k_2} is any vector in R^{k_2} that lies on the unit sphere, i.e., $||s_{k_2}|| = 1$. Note that $\xi_{1,h} = \pm \xi_{2,h}$ when $h_{11} = 0$ and $h_2 = \pm 1$.

Using (40), we show in Appendix B that for any $h \in H$,

$$\begin{pmatrix} T_n^{**}(\theta_0) \\ \widehat{\sigma}_u^2/\sigma_u^2 \end{pmatrix} \rightarrow_d \begin{pmatrix} \eta_h^{**} \\ \eta_{u,h}^2 \end{pmatrix} \tag{41}$$

under $\{\gamma_{n,h} : n \geq 1\}$, where η_h^{**} and $\eta_{u,h}^2$ are defined as follows: (i) For $h_{11} < \infty$ and $|h_{12}| < \infty$, define

$$\begin{pmatrix} \eta_h^{**} \\ \eta_{u,h}^2 \end{pmatrix} = \begin{pmatrix} (1 - T_{1,h}/T_{2,h})^{1/2} T_{3,h} \\ (1 - h_2 \xi_{1,h}/\xi_{2,h})^2 + (1 - h_2^2)\xi_{1,h}^2/\xi_{2,h}^2 \end{pmatrix}, \quad \text{where}$$

$$\begin{aligned} T_{1,h} &= (h_{12} + \psi_{Xv,h_2})^2, \\ T_{2,h} &= T_{1,h} + h_{11}^2 + \psi_{Zv,h_2}' \psi_{Zv,h_2} + 2h_{11} s_{k_2}' \psi_{Zv,h_2}, \quad \text{and} \\ T_{3,h} &= -(h_{12} + \psi_{Xv,h_2})\xi_{1,h}/\xi_{2,h} + \psi_{Xu,h_2}. \end{aligned} \tag{42}$$

Note that the random variable $\eta_{u,h}$ is positive a.s. except when $h_{11} = 0$ and $h_2 = \pm 1$. In the latter case, $\eta_{u,h} = 0$ a.s. because $\xi_{1,h} = \pm \xi_{2,h}$.

(ii) For $h_{11} < \infty$ and $|h_{12}| = \infty$, define

$$\begin{pmatrix} \eta_h^{**} \\ \eta_{u,h}^2 \end{pmatrix} = \begin{pmatrix} -(h_{11}^2 + \psi_{Zv,h_2}' \psi_{Zv,h_2} + 2h_{11} s_{k_2}' \psi_{Zv,h_2})^{1/2} \xi_{1,h}/\xi_{2,h} \\ (1 - h_2 \xi_{1,h}/\xi_{2,h})^2 + (1 - h_2^2)\xi_{1,h}^2/\xi_{2,h}^2 \end{pmatrix}. \tag{43}$$

(iii) For $h_{11} = \infty$, define $\eta_h^{**} \sim N(0, 1)$ and $\eta_{u,h}^2 = 1$ for any value of h_{12}.

Let J_h^{**} denote the distribution of η_h^{**}. The asymptotic distribution function J_h of $T_n(\theta_0)$ is given by $J_h = J_h^{**}$, $-J_h^{**}$, and $|J_h^{**}|$ for the upper, lower, and symmetric tests, respectively. Equation (41) implies that Assumption B holds for $T_n(\theta_0)$ as defined above.

We now verify Assumptions C–F for this example. We assume b is chosen such that Assumption C holds. Assumptions D and E hold by the i.i.d. assumption. Assumption G holds automatically because the subsample statistics are defined to satisfy Assumption Sub2. The distribution function of η_h^{**} is continuous and strictly increasing on R except when $h_{11} = 0$, $|h_{12}| < \infty$, and $h_2 = \pm 1$. In the latter case,

(i) $\eta_h^{**} = \mp(1 - T_{1,h}/T_{2,h})^{1/2} h_{12}$ (because $\xi_{1,h} = \pm\xi_{2,h}$ and $\psi_{Xv,h_2} = \pm\psi_{Xu,h_2}$), (ii) when $h_{12} \neq 0$, η_h^{**} has support on R_+ or R_- and has a continuous and strictly increasing distribution function on R_+ or R_-, (iii) hence, when $h_{12} \neq 0$, $J_h(c_h(1-\alpha)+\varepsilon) > J_h(c_h(1-\alpha)) = 1-\alpha$, (iv) when $h_{12} = 0$, η_h^{**} is a pointmass at zero, and (v) hence, when $h_{12} = 0$, $J_h(c_h(1-\alpha)) = 1 > 1-\alpha$. In consequence, Assumptions F and J hold for $\alpha \in (0,1)$.

6.2.3. Asymptotic Size. For upper and lower one-sided and symmetric two-sided tests and for all $\rho_U \in (0,1]$, the asymptotic sizes of the nominal level α FCV and subsampling tests are

$$AsySz(\theta_0) = \sup_{h \in H_{\rho_U}} P(\eta_h > \eta_{u,h} c_\infty(1-\alpha)) \quad \text{and}$$

$$AsySz(\theta_0) = \sup_{(g,h) \in GH_{\rho_U}} [1 - J_h(c_g(1-\alpha))], \tag{44}$$

respectively, where $\eta_h = \eta_h^{**}, -\eta_h^{**}$, and $|\eta_h^{**}|$ and $J_h = J_h^{**}, -J_h^{**}$, and $|J_h^{**}|$ for the upper, lower, and symmetric tests, respectively, $\eta_h^{**} \sim J_h^{**}$, and GH_{ρ_U} denotes the set GH defined in (19) when $H = H_{\rho_U}$. The subsampling result in (44) holds for all $\rho_U \in (0,1]$ by Theorem 1.[3] The FCV result in (44) holds for all $\rho_U \in (0,1)$ by Theorem 1. The FCV result in (44) holds for $\rho_U = 1$ because $AsySz(\theta_0) \to 1$ as $\rho_U \to 1$ (by numerical calculation) and $AsySz(\theta_0)$ for $\rho_U = 1$ is greater than or equal to $AsySz(\theta_0)$ for all $\rho_U \in (0,1)$.[4]

When $\rho_U = 1$, $AsySz(\theta_0)$ in (44) equals one for each of the subsampling tests. This holds because when $h_{11} = 0$, $|h_{12}| < \infty$, and $h_2 = \pm 1$, we have (i) $\eta_h^{**} = \mp(1 - T_{1,h}/T_{2,h})^{1/2} h_{12}$, (ii) $g_{11} = 0$, $g_{12} = 0$, $g_2 = h_2 = \pm 1$, where $g = (g_1, g_2)$, $g_1 = (g_{11}, g_{12})$, and $(g,h) \in GH_{\rho_U}$, (iii) $\eta_g^{**} = \mp(1 - T_{1,g}/T_{2,g})^{1/2} g_{12} = 0$ by part (i), (iv) $c_g(1-\alpha) = 0$ for all $\alpha \in (0,1)$, and (v) $J_h(c_g(1-\alpha)) = J_h(0) = 0$, where the second equality holds for upper one-sided, lower one-sided, and symmetric two-sided tests for all h such that $\mp h_{12} > 0$, $\mp h_{12} < 0$, and $h_{12} \neq 0$, respectively, because $(1 - T_{1,h}/T_{2,h})^{1/2} > 0$ a.s.

For the nominal level α equal-tailed subsampling test, $T_n(\theta_0) = T_n^{**}(\theta_0)$, $\eta_h^{**} \sim J_h^{**}$, $J_h = J_h^{**}$, and its asymptotic size for all $\rho_U \in (0,1]$ is

$$AsySz(\theta_0) = \sup_{(g,h) \in GH_{\rho_U}} [1 - J_h(c_g(1-\alpha/2)) + J_h(c_g(\alpha/2))]. \tag{45}$$

The result in (45) holds by Corollary 2 (by analogous arguments to those given in the endnotes above.) Furthermore, when $\rho_U = 1$, $AsySz(\theta_0)$ of the equal-tailed subsampling test equals one (using the same argument as above for one-sided and symmetric two-sided subsampling tests).

For the FCV and partially-studentized subsampling tests, Table 2 provides the maximum asymptotic rejection probabilities ($\times$ 100) for a given value of the asymptotic correlation h_2 for a range of h_2 values in $[0,1]$. (As in Table 1, these maxima are over $h \in H_{\rho_U}$ for FCV tests and $(g,h) \in GH_{\rho_U}$ for subsampling tests

TABLE 2. Weak IV example: maximum asymptotic null rejection probabilities ($\times 100$) as a function of the true correlation h_2 for nominal 5% tests

	Upper 1-sided		Lower 1-sided		Sym 2-sided		Eq-tail 2-sided	
h_2	Sub	FCV	Sub	FCV	Sub	FCV	Sub	FCV
1.00	100	100	100	100	100	100	100	100
.99	100	99.8	100	99.8	100	99.7	100	99.7
.95	98.9	97.2	100	97.3	98.3	95.9	100	95.9
.90	95.0	91.9	100	91.9	92.4	88.5	100	88.5
.80	80.6	77.3	99.6	77.4	73.7	69.3	99.1	69.3
.60	48.0	44.4	88.4	44.4	37.0	31.6	81.2	31.6
.40	24.5	18.8	55.2	18.8	16.4	9.6	42.6	9.6
.20	11.8	7.3	21.5	7.3	7.6	5.0	13.8	5.0
.00	5.2	5.1	5.3	5.1	5.3	5.0	5.3	5.0

with h_2 fixed.) Negative values of h_2 are not given because the values in Table 2 are invariant to the sign of h_2 for the FCV and symmetric two-sided subsampling test and the values for the upper one-sided subsampling test for h_2 negative equal that of $-h_2$ for the lower one-sided subsampling test and vice versa. The nominal level of the tests is $\alpha = .05$ and $k_2 = 5$ IVs are considered. The numbers in Table 2 are obtained by simulation using 50,000 simulation repetitions. When maximization over h_{11} and h_{12} is needed, a grid of size 0.02 is used for $h_{1j} \in [0, .1]$, 0.2 is used for $h_{1j} \in [.1, .9]$, 2.0 is used for $h_{1j} \in [1, 9]$, and 200 is used for $h_{1j} \in [10, 1010]$ for $j = 1, 2$.

For the FCV and symmetric and equal-tailed two-sided subsampling tests, the maximum of the values in the appropriate column in Table 2 over $h_2 \in [0, \rho_U]$ gives the asymptotic size for the parameter space $\Gamma_{\rho_U} = [-\rho_U, \rho_U]$ (up to numerical and finite grid approximations). For each of the one-sided subsampling tests, the asymptotic size for the parameter space $\Gamma_{\rho_U} = [-\rho_U, \rho_U]$ is given by the maximum of the values in the columns of Table 2 over $h_2 \in [0, \rho_U]$ for both the upper and lower subsampling tests (because the maximum rejection rates for negative values of h_2 for the upper test equal those for $-h_2$ for the lower test and vice versa).

Table 2 shows that the asymptotic sizes of the FCV and subsampling tests are all quite poor even if h_2 is bounded away from one by a substantial amount.

All of the subsampling results above are based on the partially-studentized statistic $T_n^{**}(\theta_0)$. Using the result above that $\widehat{\sigma}_u^2/\sigma_u^2 \rightarrow_d \eta_{u,h}^2$, analogous results to (44) and (45) with η_h^{**} replaced by $\eta_h^{**}/\eta_{u,h}$ can be obtained for subsampling tests based on the fully-studentized statistic $T_n^{**}(\theta_0)/\widehat{\sigma}_u$. For this statistic as well, we find that all types of subsampling tests have asymptotic size equal to one when $\rho_U = 1$. A table analogous to Table 2 but based on the fully-studentized test statistic shows that fully-studentizing the test statistic does not improve the asymptotic sizes of the subsampling tests. In particular, the values in such a table for the fully-studentized statistic (not reported) for $h_2 \geq .80$ are mostly very similar to those in

Table 2 (and always at least as large), and the values for $h_2 \leq .60$ are larger than those in Table 2.

We conclude by contrasting the negative findings of this section regarding the asymptotic size of all types of FCV and subsampling tests with the results in Andrews and Guggenberger (2005). The latter paper considers the same model and 2SLS estimator as here but inference is focused on tests concerning the coefficient on the endogenous variable, i.e., $H_0 : \beta = \beta_0$, rather than on the coefficient on an exogenous variable. For testing $H_0 : \beta = \beta_0$, the symmetric two-sided subsampling test has asymptotic size that is equal, or almost equal, to its nominal size, depending on the value of k_2. On the other hand, the FCV and one-sided and equal-tailed two-sided subsampling tests of $H_0 : \beta = \beta_0$ exhibit the same large asymptotic size distortions that are found here for tests of $H_0 : \theta = \theta_0$.

7. ASSUMPTION G

In this section we give sufficient conditions for Assumption G when Assumption Sub1 holds and $T_n(\theta_0)$ is a t statistic. The results and proof are variants of Theorems 11.3.1(i) and 12.2.2(i) and their proofs in PRW. Let $\widehat{\theta}_n$ be an estimator of a scalar parameter θ based on a sample of size n. Let $\widehat{\sigma}_n$ $(\in R)$ be an estimator of the scale of $\widehat{\theta}_n$. For alternatives of the sort (i) $H_1 : \theta > \theta_0$, (ii) $H_1 : \theta < \theta_0$, and (iii) $H_1 : \theta \neq \theta_0$, respectively, the t statistic is defined to satisfy:

Assumption t1. (i) $T_n(\theta_0) = T_n^*(\theta_0)$, or (ii) $T_n(\theta_0) = -T_n^*(\theta_0)$, or (iii) $T_n(\theta_0) = |T_n^*(\theta_0)|$, where $T_n^*(\theta_0) = \tau_n(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n$ and τ_n is some known normalization constant.

In most cases, $\tau_n = n^{1/2}$. This is true even in a unit root time series example. When θ is the lower bound of the support of a random variable, however, $\tau_n = n$.

A common case considered in the subsampling literature is when $T_n(\theta_0)$ is a *non-studentized t* statistic, see PRW. In this case, Assumption t1 and the following assumption hold.

Assumption t2. $\widehat{\sigma}_n = 1$.

We now give sufficient conditions for Assumption G when $\{\widehat{T}_{n,b,j}\}$ satisfy Assumption Sub1 and $T_n(\theta_0)$ is a nonstudentized t statistic.

Assumption H. $\tau_b/\tau_n \to 0$.

Assumption H is not very restrictive. When $\tau_n = n^{\lambda}$ for some $\lambda > 0$, it is implied by Assumption C(ii).

LEMMA 3. *Assumptions B, t1, t2, Sub1, and H imply Assumption G.*

Next, we provide sufficient conditions for Assumption G for the case when Assumption Sub1 holds and $T_n(\theta_0)$ is a studentized t statistic. Let $(\widehat{\theta}_{n,b,j}, \widehat{\sigma}_{n,b,j})$ be the subsample statistics that are defined exactly as $(\widehat{\theta}_n, \widehat{\sigma}_n)$ are defined, but are based on the jth subsample of size b. Define

$$U^{\sigma}_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(d_b \widehat{\sigma}_{n,b,j} \le x) \tag{46}$$

for a sequence of normalization constants $\{d_n : n \ge 1\}$.

The following are modified versions of Assumptions B, D, E, and H.

Assumption BB. (i) For some $r > 0$, all $h \in H$, all sequences $\{\gamma_{n,h} : n \ge 1\}$, some normalization sequences of positive constants $\{a_n : n \ge 1\}$ and $\{d_n : n \ge 1\}$, and some distribution (V_h, W_h) on R^2, $(a_n(\widehat{\theta}_n - \theta_0), d_n \widehat{\sigma}_n) \to_d (V_h, W_h)$ under $\{\gamma_{n,h} : n \ge 1\}$, (ii) $P_{\theta_0,\gamma_{n,h}}(\widehat{\sigma}_{n,b,j} > 0$ for all $j = 1, \ldots, q_n) \to 1$ under all sequences $\{\gamma_{n,h} : n \ge 1\}$ and all $h \in H$, and (iii) $W_h(0) = 0$ for all $h \in H$.

Assumption DD. (i) $\{(\widehat{\theta}_{n,b,j}, \widehat{\sigma}_{n,b,j}) : j = 1, \ldots, q_n\}$ are identically distributed under any $\gamma \in \Gamma$ for all $n \ge 1$ and (ii) $(\widehat{\theta}_{n,b,1}, \widehat{\sigma}_{n,b,1})$ and $(\widehat{\theta}_b, \widehat{\sigma}_b)$ have the same distribution under any $\gamma \in \Gamma$ for all $n \ge 1$.

Assumption EE. For all $h \in H$ and all sequences $\{\gamma_{n,h} : n \ge 1\}$ with corresponding normalization $\{d_n : n \ge 1\}$ as in Assumption BB, $U^{\sigma}_{n,b}(x) - \mathrm{E}_{\theta_0,\gamma_{n,h}} U^{\sigma}_{n,b}(x) \to_p 0$ under $\{\gamma_{n,h} : n \ge 1\}$ for all $x \in R$.

Assumption HH. $a_b/a_n \to 0$.

In most examples, the normalization sequences $\{a_n : n \ge 1\}$ and $\{d_n : n \ge 1\}$ in Assumptions BB, EE, and HH do not depend on $\{\gamma_{n,h} : n \ge 1\}$. In consequence, for notational simplicity, this dependence is suppressed. For example, in a model with i.i.d. or stationary strong mixing observations, one often takes $d_n = 1$ for all n, W_h to be a pointmass distribution with pointmass at the probability limit of $\widehat{\sigma}_n$, and $a_n = n^{1/2}$.

However, in some cases the normalization sequences $\{a_n : n \ge 1\}$ and $\{d_n : n \ge 1\}$ need to depend on $\{\gamma_{n,h} : n \ge 1\}$. For example, this occurs in an autoregressive model with a root that is less than or equal to one, see Andrews and Guggenberger (2009a). When $\{a_n : n \ge 1\}$ and $\{d_n : n \ge 1\}$ depend on $\{\gamma_{n,h} : n \ge 1\}$, it must be the case that $\tau_n = a_n(\gamma_{n,h})/d_n(\gamma_{n,h})$ does not depend on $\{\gamma_{n,h} : n \ge 1\}$. Also, in this case, Assumption HH becomes: For all sequences $\{\gamma_{n,h} : n \ge 1\}$ for which $b^r \gamma_{n,h,1} \to g_1$ for some $g_1 \in R^p_\infty$, $a_b(\gamma_{n,h})/a_n(\gamma_{n,h}) \to 0$. When d_n depends on $\gamma_{n,h}$, the normalization constant d_b that appears in $U^{\sigma}_{n,b}(x)$ in Assumption EE is $d_b = d_b(\gamma_{n,h})$.

Assumption BB implies Assumption B with $\tau_n = a_n/d_n$ (by the continuous mapping theorem using Assumption BB(iii)). Assumption DD implies

Assumption D. Assumption DD is not restrictive, given the standard methods of defining subsample statistics. Assumption EE holds automatically when the observations are i.i.d. for each $\gamma \in \Gamma$ or are stationary and strong mixing for each $\gamma \in \Gamma$ and satisfy $\sup_{\gamma \in \Gamma} \alpha_\gamma(m) \to 0$ as $m \to \infty$ (for the same reason that Assumption E holds in these cases). Assumption HH is implied by Assumption C in many examples. However, it does not hold if θ is unidentified when $\gamma_1 = 0$ (because uniformly consistent estimation of θ is not possible in this case and $a_n = 1$ in Assumption BB(i)). For example, this occurs in a model with weak instruments, see Andrews and Guggenberger (2005). (In this case, one needs to define the subsample statistics so that Assumption Sub2 holds, in which case Assumption G holds automatically.)

The following lemma generalizes Lemma 3. It does not impose Assumption t2.

LEMMA 4. *Assumptions t1, Sub1, A, BB, C, DD, EE, and HH imply Assumption G.*

Example 1 (cont.)

Assumption G follows from Lemma 4 in this example by noting that Assumptions BB and HH hold with $a_n = n^{1/2}$, $d_n = 1$, $\tau_n = n^{1/2}$, $V_h = J_h$, and W_h equal to pointmass at one (where, as above, we assume $\sigma_1^2 = 1$ without loss of generality).

8. PROOFS

The following lemmas are used in the proof of Theorem 1. (The expression $\kappa_n \to [\kappa_{1,\infty}, \kappa_{2,\infty}]$ used below is defined in Section 3.2.)

LEMMA 5. *Suppose (a) for some df's $L_n(\cdot)$ and $G_L(\cdot)$ on R, $L_n(x) \to_p G_L(x)$ for all $x \in C(G_L)$, (b) $T_n \to_d G_T$, where T_n is a scalar random variable and G_T is some distribution on R, and (c) for all $\varepsilon > 0$, $G_L(c_\infty + \varepsilon) > 1 - \alpha$, where c_∞ is the $1-\alpha$ quantile of G_L for some $\alpha \in (0,1)$. Then for $c_n := \inf\{x \in R : L_n(x) \geq 1-\alpha\}$, (i) $c_n \to_p c_\infty$ and (ii) $P(T_n \leq c_n) \to [G_T(c_\infty-), G_T(c_\infty)]$.*

Comment
If $G_T(x)$ is continuous at c_∞, then part (ii) yields $P(T_n \leq c_n) \to G_T(c_\infty)$.

LEMMA 6. *Suppose Assumptions A–G hold. Let $\{w_n : n \geq 1\}$ be any subsequence of $\{n\}$. Let $\{\gamma_{w_n} = (\gamma_{w_n,1}, \gamma_{w_n,2}, \gamma_{w_n,3}) : n \geq 1\}$ be a sequence of points in Γ that satisfies (a) $w_n^r \gamma_{w_n,1} \to h_1$ for some $h_1 \in R_\infty^p$, (b) $b_{w_n}^r \gamma_{w_n,1} \to g_1$ for some $g_1 \in R_\infty^p$, and (c) $\gamma_{w_n,2} \to h_2$ for some $h_2 \in R_\infty^q$. Let $h = (h_1, h_2)$, $g = (g_1, g_2)$, and $g_2 = h_2$. Then, we have*

(i) $(g,h) \in GH$,

(ii) $E_{\theta_0,\gamma_{w_n}} U_{w_n,b_{w_n}}(x) \to J_g(x)$ *for all* $x \in C(J_g)$,

(iii) $U_{w_n,b_{w_n}}(x) \to_p J_g(x)$ *for all* $x \in C(J_g)$ *under* $\{\gamma_{w_n} : n \geq 1\}$,
(iv) $L_{w_n,b_{w_n}}(x) \to_p J_g(x)$ *for all* $x \in C(J_g)$ *under* $\{\gamma_{w_n} : n \geq 1\}$,
(v) $c_{w_n,b_{w_n}}(1-\alpha) \to_p c_g(1-\alpha)$ *under* $\{\gamma_{w_n} : n \geq 1\}$, *and*
(vi) $P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq c_{w_n,b_{w_n}}(1-\alpha)) \to [J_h(c_g(1-\alpha)-), J_h(c_g(1-\alpha))]$.

(In Lemma 6, b_{w_n} denotes the subsample size b when the full-sample size is w_n.)

LEMMA 7. *Suppose Assumptions A–G hold. Let* $(g,h) \in GH$ *be given. Then, there is a sequence* $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) : n \geq 1\}$ *of points in* Γ *that satisfy conditions (a)–(c) of Lemma* 6 *and for this sequence parts (ii)–(vi) of Lemma* 6 *hold with* w_n *replaced by* n.

Proof of Lemma 5. For $\varepsilon > 0$ such that $c_\infty \pm \varepsilon \in C(G_L) \cap C(G_T)$, we have

$$L_n(c_\infty - \varepsilon) \to_p G_L(c_\infty - \varepsilon) < 1-\alpha \quad \text{and}$$
$$L_n(c_\infty + \varepsilon) \to_p G_L(c_\infty + \varepsilon) > 1-\alpha \tag{47}$$

by assumptions (a) and (c) and the fact that $G_L(c_\infty - \varepsilon) < 1-\alpha$ by the definition of c_∞. This and the definition of c_n yield

$$P(A_n(\varepsilon)) \to 1, \quad \text{where} \quad A_n(\varepsilon) = \{c_\infty - \varepsilon \leq c_n \leq c_\infty + \varepsilon\}. \tag{48}$$

There exists a sequence $\{\varepsilon_k > 0 : k \geq 1\}$ such that $\varepsilon_k \to 0$ as $k \to \infty$ and $c_\infty \pm \varepsilon_k \in C(G_L) \cap C(G_T)$ for all $k \geq 1$. Hence, part (i) holds.

Let $P(A,B)$ denote $P(A \cap B)$. For part (ii), using the definition of $A_n(\varepsilon)$, we have

$$P(T_n \leq c_\infty - \varepsilon, A_n(\varepsilon)) \leq P(T_n \leq c_n, A_n(\varepsilon)) \leq P(T_n \leq c_\infty + \varepsilon). \tag{49}$$

Hence,

$$\begin{aligned}
\limsup_{n\to\infty} P(T_n \leq c_n) &= \limsup_{n\to\infty} P(T_n \leq c_n, A_n(\varepsilon)) \\
&\leq \limsup_{n\to\infty} P(T_n \leq c_\infty + \varepsilon) = G_T(c_\infty + \varepsilon), \quad \text{and} \\
\liminf_{n\to\infty} P(T_n \leq c_n) &= \liminf_{n\to\infty} P(T_n \leq c_n, A_n(\varepsilon)) \\
&\geq \liminf_{n\to\infty} P(T_n \leq c_\infty - \varepsilon, A_n(\varepsilon)) = G_T(c_\infty - \varepsilon)
\end{aligned} \tag{50}$$

using assumption (b), $c_\infty \pm \varepsilon \in C(G_T)$, and (48). Given a sequence $\{\varepsilon_k : k \geq 1\}$ as above, (50) establishes part (ii). ■

Proof of Lemma 6. First, we prove part (i). We need to show that $g \in H$, $h \in H$, $g_2 = h_2$, and conditions (a)-(c) in the definition of GH hold. For

$m = 1, \ldots, p$, if $\gamma_{1,m}^{\ell} = 0$, then $g_{1,m}, h_{1,m} \in R_{+,\infty}$ by conditions (a) and (b) of the lemma. Likewise, if $\gamma_{1,m}^{u} = 0$, then $g_{1,m}, h_{1,m} \in R_{-,\infty}$. Otherwise, $g_{1,m}, h_{1,m} \in R_{\infty}$. Hence, by the definition of H_1, $g_1, h_1 \in H_1$. By condition (c) of the lemma, $h_2 \in \mathrm{cl}(\Gamma_2) = H_2$. Combining these results gives $g, h \in H$. By assumption of the lemma, $g_2 = h_2$. By conditions (a) and (b) of the lemma and Assumption C(ii), conditions (a)–(c) of *GH* hold. Hence, $(g, h) \in GH$.

Next, we prove part (ii). For notational simplicity, we drop the subscript θ_0 from $P_{\theta_0,\gamma}$ and $\mathrm{E}_{\theta_0,\gamma}$. We have

$$\begin{aligned} \mathrm{E}_{\gamma_{w_n}} U_{w_n,b_{w_n}}(x) &= q_{w_n}^{-1} \sum_{j=1}^{q_{w_n}} P_{\gamma_{w_n}}(T_{w_n,b_{w_n},j}(\theta_0) \le x) \\ &= P_{\gamma_{w_n}}(T_{w_n,b_{w_n},1}(\theta_0) \le x) = P_{\gamma_{w_n}}(T_{b_{w_n}}(\theta_0) \le x), \end{aligned} \tag{51}$$

where the first equality holds by definition of $U_{w_n,b_{w_n}}(x)$, the second equality holds by Assumption D(i), and the last equality holds by Assumption D(ii).

We now show that $P_{\gamma_{w_n}}(T_{b_{w_n}}(\theta_0) \le x) \to J_g(x)$ for all $x \in C(J_g)$ by showing that any subsequence $\{t_n\}$ of $\{w_n\}$ has a sub-subsequence $\{s_n\}$ for which $P_{\gamma_{s_n}}(T_{b_{s_n}}(\theta_0) \le x) \to J_g(x)$.

Given any subsequence $\{t_n\}$, select a sub-subsequence $\{s_n\}$ such that $\{b_{s_n}\}$ is strictly increasing. This can be done because $b_{w_n} \to \infty$ by Assumption C(i). Because $\{b_{s_n}\}$ is strictly increasing, it is a subsequence of $\{n\}$.

Below we show that Assumption B implies that for any subsequence $\{u_n\}$ of $\{n\}$ and any sequence $\{\gamma_{u_n}^* = (\gamma_{u_n,1}^*, \gamma_{u_n,2}^*, \gamma_{u_n,3}^*) \in \Gamma : n \ge 1\}$, that satisfies (a$'$) $u_n^r \gamma_{u_n,1}^* \to g_1$ and (b$'$) $\gamma_{u_n,2}^* \to g_2 \in R^q$, we have

$$P_{\gamma_{u_n}^*}(T_{u_n}(\theta_0) \le y) \to J_g(y), \tag{52}$$

for all $y \in C(J_g)$. We apply this result with $u_n = b_{s_n}$, $\gamma_{u_n}^* = \gamma_{s_n}$, and $y = x$ to obtain the desired result $P_{\gamma_{s_n}}(T_{b_{s_n}}(\theta_0) \le x) \to J_g(x)$, where (a$'$) and (b$'$) hold by Assumptions (b) and (c) on $\{\gamma_{w_n} : n \ge 1\}$.

For the proof of part (ii), it remains to show (52). Because $g \in H$, by definition of H there exists a sequence $\{\gamma_k^+ = (\gamma_{k,1}^+, \gamma_{k,2}^+, \gamma_{k,3}^+) \in \Gamma : k \ge 1\}$ such that $k^r \gamma_{k,1}^+ \to g_1$ and $\gamma_{k,2}^+ \to g_2$ as $k \to \infty$. Define a new sequence $\{\gamma_k^{**} = (\gamma_{k,1}^{**}, \gamma_{k,2}^{**}, \gamma_{k,3}^{**}) \in \Gamma : k \ge 1\}$ as follows. If $k = u_n$ set γ_k^{**} equal to $\gamma_{u_n}^*$. If $k \ne u_n$, set γ_k^{**} equal to γ_k^+. Clearly, $\gamma_k^{**} \in \Gamma$ for all $k \ge 1$ and $k^r \gamma_{k,1}^{**} \to g_1$ and $\gamma_{k,2}^{**} \to g_2$ as $k \to \infty$. Hence, $\{\gamma_k^{**} : k \ge 1\}$ is of the form $\{\gamma_{n,g} : n \ge 1\}$ and Assumption B implies that $P_{\gamma_k^{**}}(T_k(\theta_0) \le y) \to J_g(y)$ for all $y \in C(J_g)$. Because $\{u_n\}$ is a subsequence of $\{k\}$ and $\gamma_k^{**} = \gamma_{u_n}^*$ when $k = u_n$, the latter implies that $P_{\gamma_{u_n}^*}(T_{u_n}(\theta_0) \le y) \to J_g(y)$, as desired.

For part (iii) we have to show that $U_{w_n,b_{w_n}}(x) \to_p J_g(x)$ for all $x \in C(J_g)$ under $\{\gamma_{w_n} : n \ge 1\}$. Define a new sequence $\{\gamma_k^* = (\gamma_{k,1}^*, \gamma_{k,2}^*, \gamma_{k,3}^*) \in \Gamma : k \ge 1\}$ as follows: If $k = w_n$, set γ_k^* equal to γ_{w_n}. If $k \ne w_n$, for $m = 1, \ldots, p$, define

$$
\begin{aligned}
\gamma^*_{k,1,m} &= \max\{k^{-r}h_{1,m}, \gamma^{\ell}_{1,m}/2\} && \text{if } g_{1,m}=0 \ \& \ -\infty < h_{1,m} < 0\\
\gamma^*_{k,1,m} &= \min\{k^{-r}h_{1,m}, \gamma^{u}_{1,m}/2\} && \text{if } g_{1,m}=0 \ \& \ 0 < h_{1,m} < \infty\\
\gamma^*_{k,1,m} &= \max\{-k^{-2r}, \gamma^{\ell}_{1,m}/2\} && \text{if } g_{1,m}=h_{1,m}=0 \ \& \ \gamma^{\ell}_{1,m} < 0\\
\gamma^*_{k,1,m} &= \min\{k^{-2r}, \gamma^{u}_{1,m}/2\} && \text{if } g_{1,m}=h_{1,m}=0,\ \gamma^{\ell}_{1,m}=0,\\
& && \& \ \gamma^{u}_{1,m} > 0\\
\gamma^*_{k,1,m} &= \max\{-(b_k k)^{-r/2}, \gamma^{\ell}_{1,m}/2\} && \text{if } g_{1,m}=0 \ \& \ h_{1,m}=-\infty\\
\gamma^*_{k,1,m} &= \min\{(b_k k)^{-r/2}, \gamma^{u}_{1,m}/2\} && \text{if } g_{1,m}=0 \ \& \ h_{1,m}=\infty\\
\gamma^*_{k,1,m} &= \max\{b_k^{-r} g_{1,m}, \gamma^{\ell}_{1,m}/2\} && \text{if } -\infty < g_{1,m} < 0 \ \& \ h_{1,m}=-\infty\\
\gamma^*_{k,1,m} &= \min\{b_k^{-r} g_{1,m}, \gamma^{u}_{1,m}/2\} && \text{if } 0 < g_{1,m} < \infty \ \& \ h_{1,m}=\infty\\
\gamma^*_{k,1,m} &= \gamma^{\ell}_{1,m}/2 && \text{if } g_{1,m}=h_{1,m}=-\infty\\
\gamma^*_{k,1,m} &= \gamma^{u}_{1,m}/2 && \text{if } g_{1,m}=h_{1,m}=\infty, \qquad \textbf{(53)}
\end{aligned}
$$

where $\gamma^*_{k,1} = (\gamma^*_{k,1,1}, \ldots, \gamma^*_{k,1,p})'$, define $\gamma^*_{k,2} = \gamma_{w_{n_k},2}$, where $n_k = \max\{\ell \in N : w_\ell \leq k\}$, and define $\gamma^*_{k,3}$ to be any element of $\Gamma_3(\gamma^*_{k,1}, \gamma^*_{k,2})$. As defined, $\gamma^*_k \in \Gamma$ for all $k \geq 1$ using Assumption A(ii) and straightforward calculations show that $\{\gamma^*_k : k \geq 1\}$ satisfies (a)–(c) of Lemma 6 with $\{w_n\}$ replaced by $\{k\}$. By Assumption E, we know that $U_{k,b_k}(x) - E_{\theta_0,\gamma^*_k} U_{k,b_k}(x) \to_p 0$ under $\{\gamma^*_k : n \geq 1\}$ for all $x \in R$. Because for $k = w_n$, γ^*_k equals γ_{w_n}, the latter implies that $U_{w_n,b_{w_n}}(x) - E_{\theta_0,\gamma_{w_n}} U_{w_n,b_{w_n}}(x) \to_p 0$ under $\{\gamma_{w_n} : n \geq 1\}$ for all $x \in R$. Part (iii) then follows from part (ii).

To prove part (iv), we show that Assumptions A and G imply that

$$L_{w_n,b_{w_n}}(x) - U_{w_n,b_{w_n}}(x) \to_p 0 \quad \text{under} \quad \{\gamma_{w_n} : n \geq 1\} \quad \text{for all} \quad x \in C(J_g). \qquad \textbf{(54)}$$

This and part (iii) of the lemma establish part (iv). To show (54), define the same sequence $\{\gamma^*_k\}$ as in part (iii) that satisfies (a)–(c) of Lemma 6 with $\{w_n\}$ replaced by $\{k\}$. Hence, by Lemma 6(iii) with $\{w_n\}$ replaced by $\{k\}$, $U_{k,b_k}(x) \to_p J_g(x)$ as $k \to \infty$ under $\{\gamma^*_k : k \geq 1\}$ for all $x \in C(J_g)$. In consequence, because $\{\gamma^*_k : k \geq 1\}$ is of the form $\{\gamma_{n,h} : n \geq 1\}$ and satisfies $b_k^r \gamma^*_{k,1} \to g_1$, Assumption G implies that $L_{k,b_k}(x) - U_{k,b_k}(x) \to_p 0$ as $k \to \infty$ under $\{\gamma^*_k : k \geq 1\}$ for all $x \in C(J_g)$. Since $\gamma^*_k = \gamma_{w_n}$ for $k = w_n$, this implies that (54) holds.

Parts (v) and (vi) are established by applying Lemma 5 with $L_n(x) = L_{w_n,b_{w_n}}(x)$ and $T_n = T_{w_n}(\theta_0)$ and verifying the conditions of Lemma 5 using (I) part (iv), (II) $T_{w_n}(\theta_0) \to_d J_h$ under $\{\gamma_{w_n} : n \geq 1\}$ (which is verified below), and (III) Assumption F. The result of (II) holds because $\{\gamma_k^* : k \geq 1\}$ in the proof of part (iii) is of the form $\{\gamma_{n,h} : n \geq 1\}$ for h as defined in the statement of Lemma 6; this and Assumption B imply that $T_k(\theta_0) \to_d J_h$ as $k \to \infty$ under $\{\gamma_k^* : k \geq 1\}$; and the latter and $\gamma_k^* = \gamma_{w_n}$ for $k = w_n$ imply the result of (II). ■

Proof of Lemma 7. Define $\gamma_{n,1,m}$ as in (53) with n in place of k for $m = 1, \ldots, p$ and let $\gamma_{n,1} = (\gamma_{n,1,1}, \ldots, \gamma_{n,1,p})'$. Define $\{\gamma_{n,2} : n \geq 1\}$ to be any sequence of points in Γ_2 such that $\gamma_{n,2} \to h_2$ as $n \to \infty$. Let $\gamma_{n,3}$ be any element of $\Gamma_3(\gamma_{n,1}, \gamma_{n,2})$ for $n \geq 1$. Then, $\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3})$ is in Γ for all $n \geq 1$ using Assumption A. Also, using Assumption C, straightforward calculations show that $\{\gamma_n : n \geq 1\}$ satisfies conditions (a)–(c) of Lemma 6 with $w_n = n$. Hence, parts (ii)–(vi) of Lemma 6 hold with $w_n = n$ for $\{\gamma_n : n \geq 1\}$, as defined above. ■

Proof of Theorem 1. The proof of part (i) is similar to that of part (ii), but noticeably simpler because $c_{Fix}(1-\alpha)$ is a constant. Furthermore, the proof of the second result of part (ii) is quite similar to that of the first result. Hence, for brevity, we only prove the first result of part (ii).

We first show that $AsySz(\theta_0) \geq Max_{Sub}(\alpha)$. Equation (8) implies that for any sequence $\{\gamma_n \in \Gamma : n \geq 1\}$,

$$AsySz(\theta_0) \geq \limsup_{n\to\infty} [1 - P_{\theta_0,\gamma_n}(T_n(\theta_0) \leq c_{n,b}(1-\alpha))]. \tag{55}$$

In consequence, to show $AsySz(\theta_0) \geq Max_{Sub}(\alpha)$, it suffices to show that given any $(g,h) \in GH$, there exists a sequence $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\}$ such that

$$\limsup_{n\to\infty} [1 - P_{\theta_0,\gamma_n}(T_n(\theta_0) \leq c_{n,b}(1-\alpha))] \geq 1 - J_h(c_g(1-\alpha)). \tag{56}$$

The latter inequality holds by Lemma 7.

It remains to show $AsySz(\theta_0) \leq Max^-_{Sub}(\alpha)$. Let $RP_n(\gamma) = P_{\theta_0,\gamma}(T_n(\theta_0) > c_{n,b}(1-\alpha))$. Let $\{\gamma_n^* = (\gamma_{n,1}^*, \gamma_{n,2}^*, \gamma_{n,3}^*) \in \Gamma : n \geq 1\}$ be a sequence such that $\limsup_{n\to\infty} RP_n(\gamma_n^*) = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} RP_n(\gamma)$ $(= AsySz(\theta_0))$. Such a sequence always exists. Let $\{v_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} RP_{v_n}(\gamma_{v_n}^*)$ exists and equals $\limsup_{n\to\infty} RP_n(\gamma_n^*) = AsySz(\theta_0)$. Such a subsequence always exists.

Let $\gamma_{n,1,m}^*$ denote the mth component of $\gamma_{n,1}^*$ for $m = 1, \ldots, p$. Either (1) $\limsup_{n\to\infty} |v_n^r \gamma_{v_n,1,m}^*| < \infty$ or (2) $\limsup_{n\to\infty} |v_n^r \gamma_{v_n,1,m}^*| = \infty$. If (1) holds, then for some subsequence $\{w_n\}$ of $\{v_n\}$,

$$b_{w_n}^r \gamma_{w_n,1,m}^* \to 0 \quad \text{and}$$

$$w_n^r \gamma_{w_n,1,m}^* \to h_{1,m} \quad \text{for some} \quad h_{1,m} \in R. \tag{57}$$

If (2) holds, then either (2a) $\limsup_{n\to\infty} |b_{\upsilon_n}^r \gamma_{\upsilon_n,1,m}^*| < \infty$ or (2b) $\limsup_{n\to\infty} |b_{\upsilon_n}^r \gamma_{\upsilon_n,1,m}^*| = \infty$. If (2a) holds, then for some subsequence $\{w_n\}$ of $\{\upsilon_n\}$,

$$
\begin{aligned}
&b_{w_n}^r \gamma_{w_n,1,m}^* \to g_{1,m} \quad \text{for some} \quad g_{1,m} \in R \quad \text{and} \\
&w_n^r \gamma_{w_n,1,m}^* \to h_{1,m}, \quad \text{where} \quad h_{1,m} = \infty \quad \text{or} \quad -\infty \quad \text{with} \\
&sgn(h_{1,m}) = sgn(g_{1,m}).
\end{aligned} \tag{58}
$$

If (2b) holds, then for some subsequence $\{w_n\}$ of $\{\upsilon_n\}$,

$$
\begin{aligned}
&b_{w_n}^r \gamma_{w_n,1,m}^* \to g_{1,m}, \quad \text{where} \quad g_{1,m} = \infty \quad \text{or} \quad -\infty, \quad \text{and} \\
&w_n^r \gamma_{w_n,1,m}^* \to h_{1,m}, \quad \text{where} \quad h_{1,m} = \infty \quad \text{or} \quad -\infty \quad \text{with} \\
&sgn(h_{1,m}) = sgn(g_{1,m}).
\end{aligned} \tag{59}
$$

In addition, for some subsequence $\{w_n\}$ of $\{\upsilon_n\}$,

$$
\gamma_{w_n,2}^* \to h_2 \quad \text{for some} \quad h_2 \in \mathrm{cl}(\Gamma_2). \tag{60}
$$

By taking successive subsequences over the p components of $\gamma_{\upsilon_n,1}^*$ and $\gamma_{\upsilon_n,2}^*$, we find that there exists a subsequence $\{w_n\}$ of $\{\upsilon_n\}$ such that for each $m = 1, \ldots, p$, exactly one of the cases (57)–(59) applies and (60) holds. In consequence, conditions (a)–(c) of Lemma 6 hold. In addition, $\gamma_{w_n,3}^* \in \Gamma_3(\gamma_{w_n,1}^*, \gamma_{w_n,2}^*)$ for all $n \geq 1$ because $\gamma_{w_n}^* \in \Gamma$. Hence,

$$
RP_{w_n}(\gamma_{w_n}^*) \to [1 - J_h(c_g(1-\alpha)), 1 - J_h(c_g(1-\alpha)-)] \tag{61}
$$

by Lemma 6(vi). Also, $(g,h) \in GH$ by Lemma 6(i). Since $\lim_{n\to\infty} RP_{\upsilon_n}(\gamma_{\upsilon_n}^*) = AsySz(\theta_0)$ and $\{w_n\}$ is a subsequence of $\{\upsilon_n\}$, we have $\lim_{n\to\infty} RP_{w_n}(\gamma_{w_n}^*) = AsySz(\theta_0)$. This, (61), and $(g,h) \in GH$ imply that $AsySz(\theta_0) \leq Max_{Sub}^-(\alpha)$, which completes the proof of the first result of part (ii). ■

Proof of Lemma 4. Assume $U_{n,b}(x) \to_p J_g(x)$ for all $x \in C(J_g)$ under $\{\gamma_{n,h} : n \geq 1\}$ for some $g \in H$ and $h \in H$ such that $b^r \gamma_{n,h,1} \to g_1$ and $g_2 = h_2$. To show $L_{n,b}(x) - U_{n,b}(x) \to_p 0$ for all $x \in C(J_g)$ under $\{\gamma_{n,h}\}$, we use the argument in the proofs of Theorems 11.3.1(i) and 12.2.2(i) in PRW.

Define $R_n(t) := q_n^{-1} \sum_{j=1}^{q_n} 1(|\tau_b(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_{n,b,j}| \geq t)$. Using

$$
U_{n,b}(x-t) - R_n(t) \leq L_{n,b}(x) \leq U_{n,b}(x+t) + R_n(t) \tag{62}
$$

for any $t > 0$ (which holds for all versions (i)–(iii) of $T_n(\theta_0)$ in Assumption t1), the desired result follows once we establish that $R_n(t) \to_p 0$ under $\{\gamma_{n,h}\}$ for any fixed $t > 0$. By $\tau_n = a_n/d_n$, we have

$$
|\tau_b(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_{n,b,j}| \geq t \quad \text{iff} \quad (a_b/a_n) a_n |\widehat{\theta}_n - \theta_0| \geq d_b \widehat{\sigma}_{n,b,j} t, \tag{63}
$$

provided $\widehat{\sigma}_{n,b,j} > 0$, which by Assumption BB(ii) holds uniformly in $j = 1, \ldots, q_n$ wp$\to$1. (In the case where a_n and d_n depend on $\gamma_{n,h}$, the expression on the rhs of (63) is $(a_b(\gamma_{n,h})/a_n(\gamma_{n,h}))a_n(\gamma_{n,h})|\widehat{\theta}_n - \theta_0| \geq d_b(\gamma_{n,h})\widehat{\sigma}_{n,b,j}t$.) By Assumption BB(i) and HH, $(a_b/a_n)a_n|\widehat{\theta}_n - \theta_0| = o_p(1)$ under $\{\gamma_{n,h}\}$. Therefore, for any $\delta > 0$, $R_n(t) \leq q_n^{-1}\sum_{j=1}^{q_n} 1(\delta \geq d_b\widehat{\sigma}_{n,b,j}t) = U_{n,b}^{\sigma}(\delta/t)$ where the inequality holds wp$\to$1. Now, by an argument as in the proof of Lemma 6(ii) and (iii) (which uses Assumption EE but not Assumption G) applied to the statistic $d_n\widehat{\sigma}_n$ rather than $T_{w_n}(\theta_0)$, we have $U_{n,b}^{\sigma}(x) \to_p W_g(x)$ for all $x \in C(W_g)$ under $\{\gamma_{n,h}\}$, where $g \in H$ is defined as in Lemma 6 with $\{\gamma_{w_n}\}$ being equal to $\{\gamma_{n,h}\}$. Therefore, $U_{n,b}^{\sigma}(\delta/t) \to_p W_g(\delta/t)$ for $\delta/t \in C(W_g)$ under $\{\gamma_{n,h}\}$. By Assumption BB(iii), W_g does not have positive mass at zero and, hence, $W_g(\delta/t) \to 0$ as $\delta \to 0$. We can therefore establish that $R_n(t) \to_p 0$ for any $t > 0$ by letting δ go to zero such that $\delta/t \in C(W_g)$. ■

NOTES

1. We use the term "exact" size because we want to distinguish clearly between asymptotic size and finite-sample size. We do not use the term "finite-sample size" because it can be misunderstood easily to mean "a sample size n that is finite." By "exact size" we mean "size for a finite-sample size n." We note that the term "exact" is widely used as a synonym for "finite-sample" in the econometrics literature on finite-sample distribution theory.

We use the standard definition of the "size" of a test, viz., the maximum rejection probability of a test under the null hypothesis. By "asymptotic size" we mean the limit as $n \to \infty$ of the exact size of the test, i.e., the limit as $n \to \infty$ of the maximum rejection probability of a test under the null hypothesis. As defined, the term "asymptotic size" incorporates uniformity. This is natural because the term "size" by its standard definition incorporates uniformity.

2. The bootstrap typically is not asymptotically valid in a pointwise sense in these problems, whereas subsampling and the m out of n bootstrap are. However, if one is interested in the exact size for large n, then a method that is asymptotically valid in a pointwise sense, but not in a uniform sense, is not necessarily better than a method that is not asymptotically valid in either sense. The asymptotic size of the former can be worse than that of the latter. Asymptotic validity in a pointwise sense is a desirable feature, but it only gives partial information about the large sample properties of a procedure.

We note that the "problem" with subsampling and the m out of n bootstrap, alluded to in the title, is that they do not necessarily have correct asymptotic size. The "problem" is not that they necessarily have incorrect asymptotic size.

3. For $\rho_U < 1$, Theorem 1 delivers this result because continuity of J_h on R for all $h \in H_{\rho_U}$, which holds because $|h_2| < 1$, implies that $Max_{Sub}^{-}(\alpha) = Max_{Sub}(\alpha)$. For $\rho_U = 1$, it holds because $Max_{Sub}(\alpha) = 1$, see the text below, implies that $[Max_{Sub}(\alpha), Max_{Sub}^{-}(\alpha)] = \{1\} = \{\sup_{(g,h)\in GH_{\rho_U}} [1 - J_h(c_g(1-\alpha))]\}$.

4. Theorem 1 delivers the FCV result in (44) when $\rho_U \in (0,1)$ because continuity of J_h on R for all $h \in H_{\rho_U}$ implies that $Max_{Fix}^{-}(\alpha) = Max_{Fix}(\alpha)$. The FCV result in (44) does not follow from Theorem 1 when $\rho_U = 1$ because η_h and $\eta_{u,h}$ both equal zero a.s. when $h_2 = \pm 1$ and $h_{11} = h_{12} = 0$ and Assumption B does not necessarily hold for some sequences $\{\gamma_{n,h} : n \geq 1\}$.

REFERENCES

Anatolyev, S. (2004) Inference when a nuisance parameter is weakly identified under the null hypothesis. *Economic Letters* 84, 245–254.

Andrews, D.W.K. (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1383.

Andrews, D.W.K. (2000) Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68, 399–405.
Andrews, D.W.K. (2001) Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683–734.
Andrews, D.W.K. & P. Guggenberger (2010) Applications of subsampling, hybrid, and size-correction methods. *Journal of Econometrics*, forthcoming.
Andrews, D.W.K. & P. Guggenberger (2009a) Hybrid and size-corrected subsampling methods. *Econometrica* 77, 721–762.
Andrews, D.W.K. & P. Guggenberger (2009b) Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory* 25, 669–709.
Andrews, D.W.K. & P. Guggenberger (2009c) Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators. *Journal of Econometrics*, forthcoming.
Andrews, D.W.K. & J.H. Stock (2007) Inference with weak instruments. In R. Blundell, W. K. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society,* vol. III. Cambridge University Press.
Athreya, K.B. (1987) Bootstrap of the mean in the infinite variance case. *Annals of Statistics* 15, 724–731.
Bahadur, R.R. & Savage, L.J. (1956) The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics* 25, 1115–1122.
Beran, R. (1984) Bootstrap methods in statistics. *Jber. d. Dt. Math.Verein.* 86, 14–30.
Beran, R. (1997) Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics* 49, 1–24.
Beran, R. & M.S. Srivastava (1987) Correction: Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics* 15, 470–471.
Bickel, P.J. & D.A. Freedman (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* 9, 1196–1217.
Bickel, P.J., F. Götze, and W.R. van Zwet (1997) Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica* 7, 1–31.
Bretagnolle, J. (1983) Lois limites du bootstrap de certaines fonctionelles. *Annals of the Institute of H. Poincaré: Probability and Statistics* 19, 281–296.
Cavanagh, C.L., G. Elliot, & J.H. Stock (1995). Inference in models with nearly integrated regressors. *Econometric Theory* 11, 1131–1147.
Dufour, J.-M (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1387.
Dufour, J.-M. (2003) Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics* 36, 767–808.
Dümbgen, L. (1993) On differentiable functions and the bootstrap. *Probability Theory and Related Fields* 95, 125–140.
Eaton, M.L. & D.E. Tyler (1991) On Wieland's inequality and its applications to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Annals of Statistics* 19, 260–271.
Giné, E. & J. Zinn (1990) Bootstrapping general empirical measures. *Annals of Statistics* 18, 851–869.
Gleser, L.J. & J.T. Hwang (1987) The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models. *Annals of Statistics* 15, 1351–1362.
Guggenberger, P. (2007) The impact of a Hausman pretest on the asymptotic size of a hypothesis test: The panel data case. Unpublished working paper, Department of Economics, UCLA.
Guggenberger, P. (2009) On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Working paper, Department of Economics, UCLA.
Guggenberger, P. (2010) The impact of a Hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory* 26, 362–376.
Guggenberger, P. & M. Wolf (2004) Subsampling tests of parameter hypotheses and overidentifying restrictions with possible failure of identification. Unpublished working paper, UCLA.
Hahn, J. & J. Hausman (2003) Weak instruments: Diagnosis and cures in empirical economics. *American Economic Review* 93, 118–125.

Hájek, J. (1971) Limiting properties of likelihoods and inference. In V.P. Godambe & D.A. Sprott (eds.), *Foundations of Statistical Inference: Proceedings of the Symposium on the Foundations of Statistical Inference, Univ. Waterloo, Ontario,* pp. 142–159. Holt, Rinehart, and Winston.

Hall, P. & B. Jing (1995) Uniform coverage bounds for confidence intervals and Berry-Esseen theorems for Edgeworth expansion. *Annals of Statistics* 23, 363–375.

Imbens, G. & C.F. Manski (2004) Confidence intervals for partially identified parameters. *Econometrica* 72, 1845–1857.

Kabaila, P. (1995) The effect of model selection on confidence regions and prediction regions. *Econometric Theory* 11, 537–549.

Kleibergen, F. (2008) Size correct subset statistics for the linear IV regression model. Unpublished working paper, Brown University.

LeCam, L. (1953) On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *University of California Publications in Statistics* 1, 277–330.

Leeb, H. & B.M. Pötscher (2005) Model selection and inference: Facts and fiction. *Econometric Theory* 21, 21–59.

Leeb, H. & B.M. Pötscher (2006) Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results. *Econometric Theory* 22, 69–97.

Lehmann, E.L. & J.P. Romano (2005) *Testing Statistical Hypotheses*, 3rd ed. Wiley.

Loh, W.-Y. (1985) A new method for testing separate families of hypotheses. *Journal of the American Statistical Association* 80, 362–368.

Mikusheva, A. (2007) Uniform inferences in autoregressive models. *Econometrica* 75, 1411–1452.

Pfanzagl, J. (1973) The accuracy of the normal approximation for estimates of vector parameters. *Z. Wahr. verw. Geb.* 25, 171–198.

Politis, D.N. & J.P. Romano (1994) Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22, 2031–2050.

Politis, D.N., J.P. Romano, & M. Wolf (1999) *Subsampling.* Springer.

Pötscher, B.M. (2002) Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* 70, 1035–1065.

Rao, C.R. (1963) Criteria of estimation in large samples. *Sankhya* 25, 189–206.

Rao, C.R. (1973) *Linear Statistical Inference and Its Applications,* 2nd ed. Wiley.

Romano, J.P. (1989) Do bootstrap confidence procedures behave well uniformly in P? *Canadian Journal of Statistics* 17, 75–80.

Romano, J.P. & A.M. Shaikh (2005) Inference for the identified set in partially identified econometric models. Unpublished working paper, University of Chicago.

Romano, J.P. & A.M. Shaikh (2008) Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Inference and Planning (Special Issue in Honor of T.W. Anderson)* 138, 2786–2807.

Romano, J.P. & M. Wolf (2001) Subsampling intervals in autoregressive models with linear time trend. *Econometrica* 69, 1283–1314.

Samworth, R. (2003) A note on methods of restoring consistency to the bootstrap. *Biometrika* 90, 985–990.

Sen, P.K. (1979) Asymptotic properties of maximum likelihood estimators based on conditional specification. *Annals of Statistics* 7, 1019–1033.

Sen, P.K. & A.K.M.E. Saleh (1987) On preliminary test and shrinkage M-estimation in linear models. *Annals of Statistics* 15, 1580–1592.

Shao, J. (1994) Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society* 112, 1251–1262.

Shao, J. (1996) Bootstrap model selection. *Journal of the American Statistical Association* 91, 655–665.

Shao, J. & C.J.F. Wu (1989) A general theory for jackknife variance estimation. *Annals of Statistics* 15, 1563–1579.

Sheehy, A. & J.A. Wellner (1992) Uniform Donsker classes of functions. *Annals of Statistics* 20, 1983–2030.

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

Stock, J.H. (1991) Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435–459.

Stock, J.H., J.H. Wright, & M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20, 518–529.

Swanepoel, J.W.H. (1986) A note on proving that the (modified) bootstrap works. *Communications in Statistics: Theory Methods* 15, 3193–3203.

Wu, C.F.J. (1990) On the asymptotic properties of the jackknife histogram. *Annals of Statistics* 18, 1438–1452.

APPENDIX A: Tests When a Nuisance Parameter May Be Near a Boundary

Here we show that $Max^{-}_{Type}(\alpha) = Max_{Type}(\alpha)$ for $Type = Fix$ and Sub in this example and verify the formulae given in (25). For all $h = (h_1, h_2) \in H$ with $|h_2| < 1$, $\pm J^*_h(x)$ and $|J^*_h(x)|$ are continuous at all $x \in R$. If $h_2 = 1$, $J^*_h(x), -J^*_h(x)$, and $|J^*_h(x)|$ have jumps at $x = -h_1, h_1$, and h_1, respectively, but are continuous for all other $x \in R$. Likewise, if $h_2 = -1$, $J^*_h(x), -J^*_h(x)$, and $|J^*_h(x)|$ have jumps at $x = h_1, -h_1$, and h_1, respectively, but are continuous for all other $x \in R$. In addition, $J_h = J^*_h$ is stochastically increasing (decreasing) in h_1 for $h_2 < 0$ ($h_2 \geq 0$).

Using these results, for $J_h = J^*_h$, we have

$$Max^{-}_{Fix}(\alpha) = 1 - \inf_{h \in H} J_h(c_{Fix}(1-\alpha)-) = 1 - \inf_{h_2 \in [0,1]} J_{(0,h_2)}(z_{1-\alpha}) \quad \text{and}$$

$$\begin{aligned} Max^{-}_{Sub}(\alpha) &= 1 - \min\Bigg\{ \inf_{(g,h)\in GH} J_h(c_g(1-\alpha)), \inf_{((g_1,-1),(h_1,-1))\in GH} \\ &\qquad \left\{ J_{(h_1,-1)}(c_{(g_1,-1)}(1-\alpha)-)\right\}\Bigg\} \\ &= 1 - \inf_{h_2 \in [-1,0]} J_\infty(c_{(0,h_2)}(1-\alpha)). \end{aligned} \quad \textbf{(A.1)}$$

In the second and last equalities of (A.1), we use that $J_h = J^*_h$ is stochastically increasing (decreasing) in h_1 for $h_2 < 0$ ($h_2 \geq 0$), which implies that

$$\inf_{h_1 \in [0,\infty], h_2 \in [-1,0)} J_{(h_1,h_2)}(z_{1-\alpha}-) = J_\infty(z_{1-\alpha}) = 1-\alpha,$$

$$\begin{aligned} &\inf_{(g,h)\in GH, h_2 \in [0,1]} J_h(c_g(1-\alpha)) \\ &\quad = \min\Bigg\{ \inf_{h_1 \in [0,\infty), h_2 \in [0,1]} J_{(h_1,h_2)}(c_{(0,h_2)}(1-\alpha)), \\ &\qquad \inf_{h_1 \in [0,\infty], h_2 \in [0,1]} J_\infty(c_{(h_1,h_2)}(1-\alpha))\Bigg\} \end{aligned}$$

$$= \min\{1-\alpha, 1-\alpha\} = 1-\alpha, \quad \text{and}$$

$$\inf_{((g_1,-1),(h_1,-1))\in GH} J_{(h_1,-1)}(c_{(g_1,-1)}(1-\alpha)-) = J_\infty(c_{(0,-1)}(1-\alpha)). \tag{A.2}$$

By the same argument as above, $Max_{Fix}(\alpha)$ and $Max_{Sub}(\alpha)$ equal the right-hand side expressions in (A.1). This implies that $Max^-_{Type}(\alpha) = Max_{Type}(\alpha)$ for $Type = Fix$ and Sub for $J_h = J^*_h$ and verifies the expressions for $Max_{Type}(\alpha)$ given in (25).

The proof that $Max^-_{Type}(\alpha) = Max_{Type}(\alpha)$ for lower one-sided tests is the same with h_2 replaced by $-h_2$. The proof for symmetric two-sided tests is similar.

Next, we show that $Max^{\ell-}_{ET,Fix}(\alpha) = Max^{r-}_{ET,Fix}(\alpha)$. For $\alpha < 1/2$, we have

$$\begin{aligned}
&\sup_{h\in H:h_2=1} [1 - J_h(c_{Fix}(1-\alpha/2)-) + J_h(c_{Fix}(\alpha/2))] \\
&= \sup_{h_1\in[0,\infty]} [1 - J_{(h_1,1)}(z_{1-\alpha/2}) + J_{(h_1,1)}(z_{\alpha/2})] \\
&= \alpha/2 + \sup_{h_1\in[0,\infty]} J_{(h_1,1)}(z_{\alpha/2}) \\
&= \alpha/2 + J_\infty(z_{\alpha/2}) = \alpha,
\end{aligned} \tag{A.3}$$

where for the first and second equalities we use continuity of $J_{(h_1,1)}(x)$ for $x > 0$ and the fact that $J_{(h_1,1)}(x)$ for $x \geq 0$ does not depend on $h_1 \in [0,\infty]$. Similarly, because $\inf_{h\in H:h_2=-1} J_h(c_{Fix}(1-\alpha/2)-) = J_\infty(z_{1-\alpha/2}) = 1-\alpha/2$ and $J_{(h_1,-1)}(x)$ for $x \leq 0$ does not depend on $h_1 \in [0,\infty]$, we also have $\sup_{h\in H:h_2=-1}[1 - J_h(c_{Fix}(1-\alpha/2)-) + J_h(c_{Fix}(\alpha/2))] = \alpha$. Therefore, by continuity of $J_h(x) = J^*_h(x)$ for $|h_2| < 1$, it follows that $Max^{\ell-}_{ET,Fix}(\alpha) = Max^{r-}_{ET,Fix}(\alpha)$. Similar arguments yield $Max^{\ell-}_{ET,Sub}(\alpha) = Max^{r-}_{ET,Sub}(\alpha)$.

APPENDIX B: Tests on Parameters of Exogenous Variables in Weak IV Regression

First, we show that (41) holds. The limit distribution of $\widehat{\sigma}^2_u/\sigma^2_u$ has been derived in Andrews and Guggenberger (2005), hereafter AG (2005). We now derive the limit of $T^{**}_n(\theta_0)$. Using $X'P_{\overline{Z}} = X'$, we have

$$\begin{aligned}
T^{**}_n(\theta_0) &= (X'SX)^{1/2}(\widehat{\theta}_n - \theta_0)/\sigma_u \\
&= \left(X'P_{\overline{Z}}X - \frac{X'P_{\overline{Z}}y_2y_2'P_{\overline{Z}}X}{y_2'P_{\overline{Z}}y_2}\right)^{1/2}(X'X)^{-1}\Big(X'y_2(\beta - \widehat{\beta}_n) + X'u\Big)/\sigma_u \\
&= \left(1 - \frac{(X'y_2/(\sigma_v(X'X)^{1/2}))^2}{\sigma_v^{-2}y_2'P_{\overline{Z}}y_2}\right)^{1/2} \\
&\quad \times \left(\frac{X'Z\pi}{\sigma_v(X'X)^{1/2}} + \frac{(X'X)^{1/2}\phi}{\sigma_v} + \frac{X'v}{\sigma_v(X'X)^{1/2}}\right)
\end{aligned}$$

$$\times\left((\beta-\widehat{\beta}_n)\frac{\sigma_v}{\sigma_u}+\frac{X'u}{\sigma_u(X'X)^{1/2}}\right)$$
$$=(1-T_1/T_2)^{1/2}\times T_3, \tag{B.1}$$

where T_1, T_2, and T_3 are implicitly defined.

First, consider the case $h_{11}<\infty$ and $|h_{12}|<\infty$. We have

$$\begin{aligned}T_1 &= (X'y_2/(\sigma_v(X'X)^{1/2}))^2\\ &= \left(\frac{X'Z}{\sigma_v(X'X)^{1/2}}\pi+\frac{(X'X)^{1/2}}{\sigma_v}\phi+\frac{X'v}{\sigma_v(X'X)^{1/2}}\right)^2\\ &\to_d (h_{12}+\psi_{Xv,h_2})^2,\end{aligned} \tag{B.2}$$

using $(X'X)^{-1/2}X'Z\pi/\sigma_v=O_p(n^{-1/2})$, which holds because $n^{-1/2}X'Z=O_p(1)$ (since $EX_iZ_i=0$) and because $h_{11}<\infty$ implies that $n^{1/2}\pi/\sigma_v=O_p(1)$. Similarly, by multiplying out and by using $\overline{Z}'P_{\overline{Z}}=\overline{Z}'$, we obtain

$$\begin{aligned}T_2 &= y_2'P_{\overline{Z}}y_2/\sigma_v^2\\ &= \left\|\frac{(Z'Z)^{1/2}\pi}{\sigma_v}\right\|^2+2\frac{(X'X)^{1/2}\phi}{\sigma_v}\frac{X'Z\pi}{\sigma_v(X'X)^{1/2}}+\left(\frac{(X'X)^{1/2}\phi}{\sigma_v}\right)^2\\ &\quad+\frac{v'\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'v}{\sigma_v^2}+2\frac{\pi'(Z'Z)^{1/2}}{\sigma_v}\frac{(Z'Z)^{-1/2}Z'v}{\sigma_v}\\ &\quad+2\frac{\phi(X'X)^{1/2}}{\sigma_v}\frac{(X'X)^{-1/2}X'v}{\sigma_v} \qquad\qquad (B.3)\\ &\to_d h_{11}^2+h_{12}^2+\psi_{Xv,h_2}^2+\psi_{Zv,h_2}'\psi_{Zv,h_2}+2h_{11}s_{k_2}'\psi_{Zv,h_2}+2h_{12}\psi_{Xv,h_2},\end{aligned}$$

where the last line again uses $(X'X)^{-1/2}X'Z\pi/\sigma_v=O_p(n^{-1/2})$. Using this result a third time and using $(\widehat{\beta}_n-\beta)\sigma_v/\sigma_u\to_d\xi_{1,h}/\xi_{2,h}$, which is established in AG (2005), lead to

$$T_3\to_d-(h_{12}+\psi_{Xv,h_2})\xi_{1,h}/\xi_{2,h}+\psi_{Xu,h_2}. \tag{B.4}$$

Next, consider the case where $h_{11}<\infty$ and $|h_{12}|=\infty$. In this case, both T_1 and T_2 equal $((X'X)^{1/2}\phi/\sigma_v)^2(1+o_p(1))$. Multiplying out in T_1 and using again $(X'X)^{-1/2}X'Z\pi/\sigma_v=O_p(n^{-1/2})$ shows that

$$\begin{aligned}T_2-T_1 &= \left(n\gamma_{11}^2+\frac{v'\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'v}{\sigma_v^2}+2\frac{\pi'(Z'Z)^{1/2}}{\sigma_v}\frac{(Z'Z)^{-1/2}Z'v}{\sigma_v}\right.\\ &\quad\left.-\left(\frac{X'v}{\sigma_v(X'X)^{1/2}}\right)^2\right)\big(1+o_p(1)\big).\end{aligned} \tag{B.5}$$

Because $T_3 = [(X'X)^{1/2}\phi/\sigma_v](\beta - \widehat{\beta}_n)(\sigma_v/\sigma_u)(1+o_p(1))$, it follows that

$$\begin{aligned} T_n^{**}(\theta_0) &= \left(\frac{T_2 - T_1}{T_2}\right)^{1/2} T_3 = (T_2 - T_1)^{1/2}(\beta - \widehat{\beta}_n)\frac{\sigma_v}{\sigma_u}(1+o_p(1)) \\ &\to_d -(h_{11}^2 + \psi_{Xv,h_2}^2 + \psi_{Zv,h_2}'\psi_{Zv,h_2} + 2h_{11}s_{k_2}'\psi_{Zv,h_2} - \psi_{Xv,h_2}^2)^{1/2} \\ &\quad \times \xi_{1,h}/\xi_{2,h}. \end{aligned} \tag{B.6}$$

Next, consider the case $h_{11} = \infty$ and $|h_{12}| < \infty$. In this case, $T_3 \to_d \psi_{Xu,h_2}$ because $c_{1,n}(\beta - \widehat{\beta}_n)\sigma_v/\sigma_u = O_p(1)$ for $c_{1,n} = n^{1/2}\gamma_{n,h,1,1}$ by AG (2005). By assumption, $c_{1,n} \to \infty$ and $(X'X)^{-1/2}X'Z\pi/\sigma_v + (X'X)^{1/2}\phi/\sigma_v + (X'X)^{-1/2}X'v/\sigma_v = O_p(1)$. Because $T_1 = O_p(1)$ and $T_2 = c_{1,n}^2(1+o_p(1))$, we have $T_n^{**}(\theta_0) \to_d \psi_{Xu,h_2}$.

Finally, consider the case $h_{11} = \infty$ and $|h_{12}| = \infty$. Let $c_{2,n} = n^{1/2}\gamma_{n,h,1,2}$. By assumption, $c_{j,n} \to \infty$ for $j = 1,2$. We have $T_3 = ((c_{2,n} + O_p(1))/c_{1,n}) \times (c_{1,n}(\beta - \widehat{\beta}_n)) + (X'X)^{-1/2}X'u/\sigma_u$. From AG (2005) we have that $c_{1,n}(\beta - \widehat{\beta}_n)\sigma_v/\sigma_u \to_d s_{k_2}'\psi_{Zu,h_2} \sim N(0,1)$, and $s_{k_2}'\psi_{Zu,h_2}$ is independent of $\psi_{Xu,h_2} \sim N(0,1)$. We have $T_1/T_2 = c_{2,n}^2(1 + o_p(1))/(c_{1,n}^2 + c_{2,n}^2)$ and thus, asymptotically, $T_n^{**}(\theta_0)$ is distributed as $c_{1,n}(c_{1,n}^2 + c_{2,n}^2)^{-1/2}((c_{2,n}/c_{1,n})s_{k_2}'\psi_{Zu,h_2} + \psi_{Xu,h_2})$. Being the sum of two independent normal random variables, the limit distribution of $T_n^{**}(\theta_0)$ is a standard normal. This completes the verification of (41).