THE LIMIT OF FINITE-SAMPLE SIZE AND A PROBLEM WITH SUBSAMPLING

By

Donald W. K. Andrews and Patrik Guggenberger

March 2007

COWLES FOUNDATION DISCUSSION PAPER NO. 1605

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

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

The Limit of Finite-Sample Size and a Problem with Subsampling

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

Patrik Guggenberger
Department of Economics
UCLA

June 2005
Revised: March 2007

[1]Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0417911. Guggenberger gratefully acknowledges research support from a faculty research grant from UCLA in 2005. For helpful comments, we thank Victor Chernozukhov, Russell Davidson, Hannes Leeb, Azeem Shaikh, and the participants at various seminars and conferences at which the paper was presented.

Abstract

This paper considers inference based on a test statistic that has a limit distribution that is discontinuous in a nuisance parameter or the parameter of interest. The paper shows that subsample, $b_n < n$ bootstrap, and standard fixed critical value tests based on such a test statistic often have asymptotic size—defined as the limit of the finite-sample size—that is greater than the nominal level of the tests. We determine precisely the asymptotic size of such tests under a general set of high-level conditions that are relatively easy to verify. The high-level conditions are verified in several examples. Analogous results are established for confidence intervals.

The results apply to tests and confidence intervals (i) when a parameter may be near a boundary, (ii) for parameters defined by moment inequalities, (iii) based on super-efficient or shrinkage estimators, (iv) based on post-model selection estimators, (v) in scalar and vector autoregressive models with roots that may be close to unity, (vi) in models with lack of identification at some point(s) in the parameter space, such as models with weak instruments and threshold autoregressive models, (vii) in predictive regression models with nearly-integrated regressors, (viii) for non-differentiable functions of parameters, and (ix) for differentiable functions of parameters that have zero first-order derivative.

Examples (i)-(iii) are treated in this paper. Examples (i) and (iv)-(vi) are treated in sequels to this paper, Andrews and Guggenberger (2005a, b). In models with unidentified parameters that are bounded by moment inequalities, i.e., example (ii), certain subsample confidence regions are shown to have asymptotic size equal to their nominal level. In all other examples listed above, some types of subsample procedures do not have asymptotic size equal to their nominal level.

Keywords: Asymptotic size, $b < n$ bootstrap, finite-sample size, over-rejection, size correction, subsample confidence interval, subsample test.

JEL Classification Numbers: C12, C15.

1 Introduction

The topic of this paper is subsampling. Subsampling is a very general method for carrying out inference in econometric and statistical models, see Politis and Romano (1994). Also see Shao and Wu (1989), Wu (1990), Sherman and Carlstein (1996), and Politis, Romano, and Wolf (1999) (hereafter PRW).[2] Minimal conditions are needed for subsample tests and confidence intervals (CIs) to have desirable asymptotic properties, such as asymptotically correct rejection rates and coverage probabilities under standard asymptotics based on a fixed true probability distribution for the observations, see PRW. On the other hand, subsample methods have the disadvantage of not providing as good approximations in regular models as other methods, such as standard fixed critical value (FCV) methods based on first-order asymptotics and bootstrap procedures. In consequence, subsample methods are most useful in models that are non-regular (in the sense that test statistics of interest do not have asymptotic normal or chi-square distributions and the bootstrap is inconsistent).

This paper deals with the properties of subsampling in a broad class of non-regular models. In particular, it considers cases in which a test statistic has a discontinuity in its asymptotic distribution as a function of the true distribution that generates the observations. Numerous problems in econometrics and other areas of statistics exhibit this feature. For such problems, standard FCV procedures and bootstrap procedures typically do not provide asymptotically valid inference. In consequence, for such problems, subsample and $b < n$ bootstrap methods (where b is the bootstrap sample size) often have been advocated.

In this paper, we show that if a sequence of test statistics has an asymptotic null distribution that is discontinuous in a nuisance parameter, then a subsample test based on the test statistic does not necessarily yield the desired asymptotic level. Specifically, the limit of the finite-sample size of the test can exceed its nominal level. The same is shown to be true for a $b < n$ bootstrap test and a standard fixed critical value (FCV) test. Analogous potential problems arise with confidence sets based on subsample, $b < n$ bootstrap, and FCV methods. We note that the potential problem is not just a small sample problem—it arises with all sample sizes.

The intuition for the result stated above for a subsample test is roughly as follows. Suppose for a parameter θ we are interested in testing $H_0 : \theta = \theta_0$, a nuisance parameter γ appears under the null hypothesis, and the asymptotic distribution of the test statistic of interest is discontinuous at $\gamma = 0$. Then, a subsample test statistic based on a subsample of size $b_n << n$ behaves like it is closer to the discontinuity point $\gamma = 0$ than does the full-sample test statistic. This occurs because the variability of the subsample statistic is greater than that of the full-sample statistic and, hence, its behavior at a fixed value $\gamma \neq 0$ is harder to distinguish from its behavior at $\gamma = 0$. In consequence, the subsample statistic can have a distribution that is close to the asymptotic distribution for $\gamma = 0$, whereas the full-sample statistic has a distribution

[2] Shao and Wu (1989) and Wu (1990) refer to subsampling as the delete d jackknife.

that is close to the asymptotic distribution for $\gamma \neq 0$. If the asymptotic distribution of the test statistic for $\gamma \neq 0$ is more disperse than for $\gamma = 0$, then the subsample critical value is too small and the subsample test over-rejects the null hypothesis. On the other hand, if the asymptotic distribution of the test statistic for $\gamma \neq 0$ is less disperse than for $\gamma = 0$, then the subsample critical value is too large and the subsample test is not asymptotically similar. In fact, the limit of the finite-sample size of a subsample test depends on the whole range of behavior of the test statistic and the subsample statistic for parameter values close to $\gamma = 0$.

The intuition laid out in the previous paragraph is made rigorous by considering the behavior of subsample tests under asymptotics in which the true parameter, γ_n, drifts to the point of discontinuity $\gamma = 0$ as $n \to \infty$. Since the finite-sample size of a test is based on the supremum of the null rejection rate over all parameter values γ for given n, the limit of the finite-sample size of a test is always greater than or equal to its limit under a drifting sequence $\{\gamma_n : n \geq 1\}$. Hence, if the limit of the null rejection rate under a drifting sequence exceeds the nominal level, then the limit of the exact finite-sample null rejection rate exceeds the nominal level. Analogously, if the limit of the null rejection rate under a drifting sequence is less than the nominal level, then the test is not asymptotically similar.

We show that there are two different rates of drift such that over-rejection or under-rejection can occur. The first rate is one under which the full-sample test statistic has an asymptotic distribution that depends on a localization parameter, h, and the subsample critical values behave like the critical value from the asymptotic distribution of the statistic under $\gamma = 0$. (Under such parameter drifts, the distribution of the data typically is contiguous to their distribution under $\gamma = 0$.) The second rate is one under which the full-sample test statistic has an asymptotic distribution that is the same as for fixed $\gamma \neq 0$ and the subsample critical values behave like the critical value from the asymptotic distribution of the full-sample statistic under a drifting sequence with localization parameter h. (Under this second type of parameter drift, the distribution of the data typically is not contiguous to their distribution under $\gamma = 0$.)

The paper shows that sequences of these two types determine the limit of the finite-sample size of the test. In particular, we obtain an explicit expression for the limit of the finite-sample size of the test. This yields necessary and sufficient conditions for the limit to exceed the nominal level of the test.

The paper gives corresponding results for standard tests that are based on fixed critical values. The asymptotic results given here for subsample tests also apply to $b < n$ bootstrap tests applied to i.i.d. observations provided $b^2/n \to 0$. The reason is that subsampling based on subsamples of size b can be viewed as bootstrapping without replacement, which is not too different from bootstrapping with replacement when b^2/n is small.[3] The subsample results apply to both i.i.d. and time series observations,

[3] In an i.i.d. scenario, the distribution of a subsample of size b is the same as the conditional distribution of a nonparametric bootstrap sample of size b conditional on there being no duplicates of observations in the bootstrap sample. If $b^2/n \to 0$, then the probability of no duplicates goes to one as $n \to \infty$, see PRW, p. 48. In consequence, $b < n$ bootstrap tests and subsample tests have the same

whereas the $b < n$ bootstrap results apply only to i.i.d. observations.

The potential problems of subsample, $b < n$ bootstrap, and FCV tests outlined above carry over with some adjustments to confidence intervals (CIs) by the usual duality between tests and CIs. Some adjustments are needed because the limit of the finite-sample level of a CI depends on uniformity over $\theta \in \Theta$ and $\gamma \in \Gamma$, where Θ and Γ are the parameter spaces of θ and γ, respectively, whereas the limit of the finite-sample size of a test of $H_0 : \theta = \theta_0$ only depends on uniformity over $\gamma \in \Gamma$ for fixed θ_0.

In examples considered in the paper, the asymptotic sizes of subsample 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. For example, a subsample CI of nominal level $1 - \alpha$ based on a post-model-selection/super-efficient estimator is found to have asymptotic level of zero. In a model with a nuisance parameter near a boundary, lower one-sided, upper one-sided, symmetric two-sided, and equal-tailed two-sided subsample tests with nominal level .05 are found to have asymptotic sizes of (about) .50, .50, .10, and .525, respectively. In models with unidentified parameters that are bounded by moment inequalities, certain subsample confidence regions are shown to have asymptotic size equal to their nominal level. In an autoregressive model with an intercept and time trend, equal-tailed and symmetric two-sided subsample CIs of nominal level .95 are found to have asymptotic sizes of (about) .25 and .95, respectively, see Andrews and Guggenberger (2005a).

The results above show that a subsample test can have an asymptotic null rejection rate that equals its nominal level under any fixed true distribution, but still the limit of its finite-sample size can be greater than its nominal level. This is due to a lack of uniformity in the pointwise asymptotics. In the context of subsampling and the $b < n$ bootstrap, 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 Andrews (2000) in the context of problems due to a parameter being near a boundary, Mikusheva (2005) and Andrews and Guggenberger (2005a) in the context of first-order autoregressive models with a root that may be near unity, and Romano and Shaikh (2005a,b) in the context of inference based on moment inequalities—also see Section 9 below regarding this case. Beran (1997, p. 15) mentions that the pointwise $b < n$ bootstrap convergence typically is not locally uniform at parameter points that are not locally asymptotically equivariant, but does not discuss the consequences.

On the other hand, problems arising from lack of uniformity in asymptotics have long been recognized in the wider statistical literature. For example, Hodge's super-efficient estimator was used to show that uniformity is important in the context of asymptotic efficiency results, see LeCam (1953). The importance of uniformity to enable asymptotic results to provide good finite-sample approximations has been pointed out in many papers including Rao (1963, 1973), Hájek (1971), Pfanzagl (1973), Loh (1985), and Kabaila (1995). Related references include Bahadur and Savage (1956),

first-order asymptotic properties.

Gleser and Hwang (1987), Hall and Jing (1995), Dufour (1997), and Pötscher (2002). Romano (1989) addresses uniformity issues for some bootstrap procedures.

A sampling of references in the literature where issues of uniformity arise in applications to which the results of this paper apply is as follows: (i) for inference based on super-efficient and shrinkage estimators, see Sen and Saleh (1987), Kabaila (1995), and Leeb and Pötscher (2006), (ii) for inference based on moment inequalities, see Imbens and Manski (2004) and Romano and Shaikh (2005a), (iii) for post-model selection inference, see Sen (1979), Sen and Saleh (1987), Kabaila (1995), Leeb and Pötscher (2005), and additional references cited therein, (iv) for autoregressive models with roots that may be near unity, see Stock (1991), Andrews (1993), and Mikusheva (2005), (v) for predictive regressions with nearly integrated regressors, see Cavanagh, Elliot, and Stock (1995), (vi) for weak instruments, see Staiger and Stock (1997), Dufour (1997), and papers referenced in Andrews and Stock (2006), and (vii) for inference in threshold autoregressive models, see Anatolyev (2004). Other applications that are covered by the results of this paper for which uniformity (of the type discussed in this paper) does not seem to have been discussed explicitly in the literature include (i) tests of Granger causality in vector autoregressive models, see Choi (2005), (ii) inference for non-differentiable functions of parameters, which includes inference concerning the eigenvalues of a variance matrix, see Beran and Srivastava (1985, 1987), Eaton and Tyler (1991), Dümbgen (1993), and Shao (1994), (iii) inference for differentiable functions of parameters with zero first-order derivative, see Babu (1984) and Shao (1994), (iv) tests of stochastic dominance for random variables with finite support, see Linton, Maasoumi, and Whang (2005), and (v) one-sided Kolmogorov-Smirnov tests of incomplete models for random variables with finite support, see Galichon and Henry (2006). To treat cases (iv) and (v) for random variables with infinite support, the results of this paper need to be extended to allow the parameter h_1, defined below, to be infinite dimensional. This is a topic of future research.

The $b < n$ bootstrap has been considered in a variety of different non-regular cases including many of those listed above, e.g., see Bretagnolle (1983), Shao (1994, 1996), Beran (1997), Bickel, Götze, and van Zwet (1997), and Andrews (2000).

The results in the paper apply to some non-regular cases where the limit distribution of a test statistic is "continuous" in a nuisance parameter. For such cases, sufficient conditions are given under which subsample, $b < n$ bootstrap, and FCV tests and CIs have asymptotic levels equal to their nominal levels. To the best of our knowledge, results of this sort are not available in the literature. An example is the case where the parameter of interest is the lower bound on the support of a random variable. This example is treated in Appendix B. The bootstrap (typically) is not consistent in this case, see Bickel and Freedman (1981) and Loh (1984). Asymptotic results concerning the $b < n$ bootstrap for this example are pointwise results, see Shao (1994) and Bickel, Götze, and van Zwet (1997). This example is a special case of some production frontier and auction models that are of interest in econometrics, see Hirano and Porter (2003) and Chernozhukov and Hong (2004).

The present paper considers general test statistics, including t, likelihood ratio (LR), and Lagrange multiplier (LM) statistics. The results cover one-sided, symmetric two-sided, and equal-tailed two-sided t tests and corresponding confidence intervals. The t statistics may be studentized (i.e., of the form $\tau_n(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n$ for an estimator $\widehat{\theta}_n$, a scale estimator $\widehat{\sigma}_n$, and a normalization factor τ_n) or non-studentized (i.e., of the form $\tau_n(\widehat{\theta}_n - \theta_0)$). Non-studentized t statistics are often considered in the subsample literature, see PRW. But, studentized t statistics are needed in certain testing situations in which non-studentized statistics have rates of convergence that are parameter dependent. This occurs with unit root tests, see Romano and Wolf (2001), and with tests in the presence of weak instruments, see Guggenberger and Wolf (2004).

The main results given in the paper employ high-level assumptions. These assumptions are verified in examples concerning (i) a test with a nuisance parameter that may be near a boundary, (ii) a CI based on a post-"consistent"-model-selection/super-efficient estimator, (iii) a confidence set for an unidentified parameter that is bounded by moment inequalities, as in Chernozhukov, Hong, and Tamer (2002), Romano and Shaikh (2005a), Rosen (2005), and Soares (2005), and (iv) a CI for the lower end-point of the support of a distribution. In addition, Andrews and Guggenberger (2005a) verify the high-level conditions in examples concerning (v) tests and CIs based on a post-"conservative"-model-selection estimator and (vi) a CI for an autoregressive parameter that may be near unity. Andrews and Guggenberger (2005b) do likewise in examples concerning (vii) a test in an instrumental variables (IVs) regression model with IVs that may be weak and (viii) a CI where the parameter of interest may be near a boundary.

Andrews and Guggenberger (2005a) utilize the results of this paper to introduce and analyze various new procedures including (i) hybrid subsample/FCV, (ii) size-corrected FCV, (iii) size-corrected subsample, and (iv) size-corrected hybrid tests and CIs (and analogous $b < n$ bootstrap procedures). These procedures extend the applicability of subsample, $b < n$ bootstrap, and FCV methods to a wide variety of models whose asymptotic distributions are discontinuous in some parameter.

The remainder of the paper is organized as follows. Section 2 illustrates the basic problem with subsampling in a simple parameter near-a-boundary example. Section 3 describes the basic set-up for tests. Section 4 conveys the main ideas and results of the paper in a simple model. Although the assumptions of this model are too restrictive for many applications, the ideas and results are much easier to comprehend than in a more general framework. Sections 5 and 6 specify the general assumptions and asymptotic results of the paper for one-sided and symmetric two-sided tests. Sections 7 and 8 extend the results to equal-tailed two-sided tests and CIs, respectively. Section 9 provides several examples of the general results. Appendix A gives sufficient conditions for a technical high-level assumption that arises with studentized t statistics, provides verification of assumptions in several examples, provides proofs of the results stated in the paper, and provides an improvement of the main results of the text that is useful in some examples. Appendix B provides a proof of a simplified version of the main result of the paper. This proof is more transparent than the proof of the general result given in the paper. Appendix B also provides an example in which the parameter of interest is the lower bound on the support of a random variable.

Throughout the paper $\alpha \in (0,1)$ denotes a given constant.

2 Illustration of a Problem with Subsampling

Here we illustrate a potential problem with subsampling using a simple boundary example. (The same problem arises with the $b < n$ bootstrap.) In this example, a parameter θ_0 is restricted to be non-negative. Suppose $X_i \sim$ i.i.d. $N(\theta_0, 1)$ for $i = 1, ..., n$ 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 $\widehat{\theta}_n$ is

$$\widehat{\theta}_n \sim \max\{Z_n, 0\}, \text{ where } Z_n \sim N(\theta_0, n^{-1}). \tag{2.1}$$

The jth subsample estimator based on a subsample of size $b_n = o(n)$ is $\widehat{\theta}_{b_n,j} = \max\{\overline{X}_{b_n,j}, 0\}$, where $\overline{X}_{b_n,j} = b_n^{-1} \sum_{i=j}^{j+b_n-1} X_i$. Its distribution is

$$\widehat{\theta}_{b_n,j} \sim \max\{Z_{b_n}, 0\}, \text{ where } Z_{b_n} \sim N(\theta_0, b_n^{-1}). \tag{2.2}$$

Figure 1 shows the densities of Z_n and Z_{b_n} for the case where $\theta_0 = .15$, $n = 100$, $b_n/n = 1/10$. The estimators $\widehat{\theta}_n$ and $\widehat{\theta}_{b_n,j}$ have densities for $x > 0$ that are given by the peaked and flatter curves, respectively. The estimators equal zero with probabilities given by the areas under the peaked and flatter curves for $x < 0$. It is clear that the probability that $\widehat{\theta}_n = 0$ is much smaller than the probability that $\widehat{\theta}_{b_n,j} = 0$. In consequence, the distribution of $\widehat{\theta}_{b_n,j}$ does not properly mimic the distribution of $\widehat{\theta}_n$. The reason this occurs is that the subsample estimator behaves as though it is closer to the boundary of the parameter space than the full sample estimator because it is much more variable. In consequence, the effect of the boundary on the subsample estimator is noticeably larger than on the full-sample estimator.

Obviously, a subsample estimator has a different scale than a full-sample estimator. In consequence, one uses the distribution of a re-centered and re-scaled subsample estimator to approximate the corresponding distribution of the re-centered and re-scaled full-sample estimator. Although suggestive, Figure 1 does not make this comparison completely clear. In the present example, the re-centered and re-scaled full-sample and subsample estimators are

$$\begin{aligned} T_n = n^{1/2}(\widehat{\theta}_n - \theta_0) &= n^{1/2}(\max\{\overline{X}_n, 0\} - \theta_0) \\ &= \max\{n^{1/2}(\overline{X}_n - \theta_0), -n^{1/2}\theta_0\} \\ &\sim \max\{Z, -h\}, \end{aligned} \tag{2.3}$$

where $Z \sim N(0,1)$ and $h = n^{1/2}\theta_0$, and

$$T_{b_n,j} = b_n^{1/2}(\widehat{\theta}_{b_n,j} - \theta_0) = b_n^{1/2}(\max\{\overline{X}_{b_n,j}, 0\} - \theta_0)$$

$$= \max\{b_n^{1/2}(\overline{X}_{b_n,j} - \theta_0), -b_n^{1/2}\theta_0\}$$
$$\sim \max\{Z, \ -(b_n/n)^{1/2}\,h\}. \quad (2.4)$$

The densities of T_n and $T_{b_n,j}$ are graphed in Figure 2 for $h = 1.5$ and $b_n/n = 1/10$. The vertical lines at -1.5 and -0.5 show the probabilities that T_n and $T_{b_n,j}$ take on these two values. Clearly, the subsample distribution gives a very good approximation of the full-sample distribution in the right tail, but a poor one in the left-tail. Hence, an upper one-sided subsample CI for θ_0, which relies on a subsample critical value from the right tail of the subsample distribution, will perform well. But a subsample lower one-sided CI will perform poorly. Furthermore, equal-tailed and symmetric two-sided subsample CIs will perform poorly.

The finite-sample problem illustrated in Figure 2 is maintained in the limit as $n \to \infty$ provided $h = n^{1/2}\theta_0$ is a constant for all n (or converges to a constant). Hence, for asymptotic results to properly capture the finite-sample behavior of the subsample method when θ_0 is near the boundary, one needs to consider true values θ_0 that drift to zero as $n \to \infty$:

$$\theta_0 = \theta_{0,n} = h/n^{1/2}. \quad (2.5)$$

Such sequences capture the non-uniform convergence of the statistics T_n and $T_{b_n,j}$. (Results below show that one also has to consider sequences that drift to zero at a slower rate than $1/n^{1/2}$, i.e., non-contiguous sequences, in order to cover cases in which T_n behaves as though it is relatively far from the boundary but $T_{b_n,j}$ behaves as though it is near the boundary.) Asymptotic results that only consider fixed θ_0 values are misleading because they fail to reveal the effects of non-uniform convergence. The formal set-up below allows for the case where the parameter drifts to zero. Often, the limit of the finite-sample size of a subsample test or CI is determined by its behavior along sequences of this type.

A complete treatment of the boundary example considered above, generalized to regression models, is given in Andrews and Guggenberger (2005b) based on the results of this paper.

3 Basic Testing Set-up

We are interested in tests concerning a parameter $\theta \in R^d$ in the presence of a nuisance parameter $\gamma \in \Gamma$. Often $d = 1$, but the results allow for $d > 1$. The null hypothesis of interest is $H_0 : \theta = \theta_0$. The alternative hypothesis of interest may be one-sided or two-sided.

3.1 Test Statistic

Let $T_n(\theta_0)$ denote a test statistic based on a sample of size n for testing $H_0 : \theta = \theta_0$ for some $\theta_0 \in R^d$. The leading case that we consider is when $T_n(\theta_0)$ is a t statistic, but the results also allow $T_n(\theta_0)$ to be an LR, LM, or some other statistic. Large values of

$T_n(\theta_0)$ indicate evidence against the null hypothesis, so a test based on $T_n(\theta_0)$ rejects the null hypothesis when $T_n(\theta_0)$ exceeds some critical value.

When $T_n(\theta_0)$ is a t statistic, it is defined as follows. 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 as follows:

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 many cases, $\tau_n = n^{1/2}$. For example, this is true in boundary examples and even in a unit root example. Note that τ_n is not uniquely defined because $\widehat{\sigma}_n$ could be scaled up or down to counteract changes in the scale of τ_n. In practice this is usually not an issue because typically there is a natural definition for $\widehat{\sigma}_n$, which determines its scale.

A common case considered in the subsample 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. $\widehat{\sigma}_n = 1$.

There are cases, however, where a non-studentized test statistic has an asymptotic null distribution with a normalization factor τ_n that depends on a nuisance parameter γ. This causes problems for the standard theory concerning subsample methods, see PRW, Ch. 8. In such cases, a studentized test statistic often has the desirable property that the normalization factor τ_n does not depend on the nuisance parameter γ. This occurs with tests concerning unit roots in time series, see Romano and Wolf (2001), and with tests in the presence of weak instruments, see Guggenberger and Wolf (2004). The set-up that we consider allows for both non-studentized and studentized test statistics. Note that under Assumption t2 the order of magnitude of τ_n is uniquely determined.

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. Without loss of generality, we take the point(s) of discontinuity to be γ values for which some subvector of γ is 0.

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 $\{X_i \in R^2 : i \leq n\}$ are i.i.d. with distribution F,

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}, \; E_F X_i = \begin{pmatrix} \theta \\ \mu \end{pmatrix}, \text{ and } Var_F(X_i) = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}. \tag{3.1}$$

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 $|T_n^*(\theta_0)|$, $T_n^*(\theta_0) = \tau_n(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_{n1}$ and $\tau_n = n^{1/2}$.

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. Define $(\widehat{\theta}_n, \widehat{\mu}_n)$ to be the Gaussian quasi-ML estimator of (θ, μ) under the restriction that $\widehat{\mu}_n \geq 0$ and under the assumption that the standard deviations and correlation of X_i 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 $\widehat{\mu}_n \geq 0$. 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

$$\begin{aligned} \widehat{\theta}_n &= \overline{X}_{n1} - (\widehat{\rho}_n \widehat{\sigma}_{n1}) \min(0, \overline{X}_{n2}/\widehat{\sigma}_{n2}), \text{ where} \\ \overline{X}_{nj} &= n^{-1} \textstyle\sum_{i=1}^{n} X_{ij} \text{ for } j = 1, 2. \end{aligned} \tag{3.2}$$

3.2 Fixed Critical Values

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

$$T_n(\theta_0) > c_{Fix}(1-\alpha). \tag{3.3}$$

The results below allow $c_{Fix}(1-\alpha)$ to be any constant. However, if the discontinuity (or discontinuities) of the asymptotic null distribution of $T_n(\theta_0)$ is (are) not taken into account, one typically defines

$$c_{Fix}(1-\alpha) = c_\infty(1-\alpha), \tag{3.4}$$

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 not a point of discontinuity. For example, for studentized tests when Assumption t1(i), (ii), or (iii) holds, $c_\infty(1-\alpha)$ typically equals $z_{1-\alpha}$, $z_{1-\alpha}$, or $z_{1-\alpha/2}$, respectively, where $z_{1-\alpha}$ denotes the $1-\alpha$ quantile of the standard normal distribution. If $T_n(\theta_0)$ is an LR, LM, or Wald statistic, then $c_\infty(1-\alpha)$ typically equals the $1-\alpha$ quantile of a χ^2_d distribution, denoted $\chi^2_d(1-\alpha)$.

On the other hand, if a discontinuity at $\gamma = h^0$ is recognized, one might take the FCV to be

$$c_{Fix}(1-\alpha) = \max\{c_\infty(1-\alpha), c_{h^0}(1-\alpha)\}, \tag{3.5}$$

where $c_{h^0}(1-\alpha)$ denotes the $1-\alpha$ quantile of J_{h^0} and J_{h^0} is the asymptotic null distribution of $T_n(\theta_0)$ when $\gamma = h^0$. The FCV test based on this FCV is not likely to be asymptotically similar, but one might hope that it has asymptotic level α. The results given below show that often the latter is not true.

Example 1 (cont.). The FCVs employed in this example are the usual standard normal critical values that ignore the fact that μ may be on 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.

3.3 Subsample Critical Values

The second type of critical value that we consider is a subsample critical value. Let $\{b_n : n \geq 1\}$ be a sequence of subsample sizes. For brevity, we sometimes write b_n as b. Let $\{\widehat{T}_{n,b,j} : j = 1, ..., q_n\}$ be certain subsample statistics that are based primarily on subsamples of size b_n rather than the full sample. For example, with i.i.d. observations, there are $q_n = n!/((n - b_n)!b_n!)$ different subsamples of size b_n and $\widehat{T}_{n,b,j}$ is determined primarily by the observations in the jth such subsample. With time series observations, say $\{X_1, ..., X_n\}$, there are $q_n = n - b_n + 1$ subsamples of b_n consecutive observations, e.g., $Y_j = \{X_j, ..., X_{j+b_n-1}\}$, and $\widehat{T}_{n,b,j}$ is determined primarily by the observations in the jth subsample Y_j.

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 $\{\widehat{T}_{n,b,j} : j = 1, ..., q_n\}$. They are defined by

$$\begin{aligned} L_{n,b}(x) &= q_n^{-1} \sum_{j=1}^{q_n} 1(\widehat{T}_{n,b,j} \leq x) \text{ for } x \in R \text{ and} \\ c_{n,b}(1-\alpha) &= \inf\{x \in R : L_{n,b}(x) \geq 1 - \alpha\}. \end{aligned} \tag{3.6}$$

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

$$T_n(\theta_0) > c_{n,b}(1 - \alpha). \tag{3.7}$$

We now describe the subsample statistics $\{\widehat{T}_{n,b,j} : j = 1, ..., q_n\}$ in more detail. Let $\{T_{n,b,j}(\theta_0) : j = 1, ..., q_n\}$ be subsample statistics that are defined exactly as $T_n(\theta_0)$ is defined, but based on subsamples of size b_n rather than the full sample. For example, suppose Assumption t1 holds. Let $(\widehat{\theta}_{n,b,j}, \widehat{\sigma}_{n,b,j})$ denote the estimators $(\widehat{\theta}_b, \widehat{\sigma}_b)$ applied to the jth subsample. In this case, we have

$$\begin{aligned} \text{(i) } T_{n,b,j}(\theta_0) &= \tau_b(\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}, \text{ or} \\ \text{(ii) } T_{n,b,j}(\theta_0) &= -\tau_b(\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}, \text{ or} \\ \text{(iii) } T_{n,b,j}(\theta_0) &= |\tau_b(\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}|. \end{aligned} \tag{3.8}$$

Below we make use of the empirical distribution of $\{T_{n,b,j}(\theta_0) : j = 1, ..., q_n\}$ defined by

$$U_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta_0) \leq x). \tag{3.9}$$

In most cases, subsample critical values are based on a simple adjustment to the statistics $\{T_{n,b,j}(\theta_0) : j = 1, ..., q_n\}$, where the adjustment is designed to yield subsample statistics that behave similarly under the null and the alternative hypotheses. In particular, $\{\widehat{T}_{n,b,j} : j = 1, ..., q_n\}$ often are defined to satisfy the following condition.

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

The estimator $\widehat{\theta}_n$ is usually chosen to be a consistent estimator of θ whether or not the null hypothesis holds. Assumption Sub1 can be applied to t statistics as well as to LR and LM statistics, among others.

If consistent estimation of θ is not possible at the point of discontinuity, say when $\gamma = h^0$, as occurs when θ is not identified when $\gamma = h^0$, then taking $\{\widehat{T}_{n,b_n,j}\}$ to satisfy Assumption Sub1 is not desirable because $\widehat{\theta}_n$ is not necessarily close to θ_0 when γ is close to γ^0. For example, this occurs in the weak IV example, see Guggenberger and Wolf (2004). In such cases, it is preferable to take $\{\widehat{T}_{n,b_n,j}\}$ to satisfy the following assumption.[4]

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

The results given below for subsample tests allow for subsample statistics $\{\widehat{T}_{n,b,j}\}$ that satisfy Assumption Sub1 or Sub2 or are defined in some other way.

Example 1 (cont.). The subsample critical values in this example 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.)

3.4 Asymptotic Size

The exact size, $ExSz_n(\theta_0)$, of an FCV or subsample test is the supremum over $\gamma \in \Gamma$ of the null rejection probability under γ:

$$\begin{aligned} ExSz_n(\theta_0) &= \sup_{\gamma \in \Gamma} RP_n(\theta_0, \gamma), \text{ where } RP_n(\theta_0,\gamma) = P_{\theta_0,\gamma}(T_n(\theta_0) > c_{1-\alpha}), \\ c_{1-\alpha} &= c_{Fix}(1-\alpha) \text{ or } c_{1-\alpha} = c_{n,b}(1-\alpha), \end{aligned} \tag{3.10}$$

and $P_{\theta,\gamma}(\cdot)$ denotes probability when the true parameters are (θ, γ).[5]

We are interested in the "asymptotic size" of the test defined by

$$AsySz(\theta_0) = \limsup_{n\to\infty} ExSz_n(\theta_0). \tag{3.11}$$

This definition should not be controversial. Our interest is in the exact finite-sample size of the test. We use asymptotics to approximate this. 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.[6]

If $AsySz(\theta_0) > \alpha$, then the nominal level α test has asymptotic size greater than α and the test does not have correct asymptotic level.

---

[4] When Assumption t1 holds, subsample statistics $\{\widehat{T}_{n,b_n,j}\}$ that satisfy Assumption Sub2 typically yield nontrivial power because the normalization constant τ_n satisfies $\tau_{b_n}/\tau_n \to 0$.

[5] We remind the reader that the *size* of a test is equal to the supremum of its rejection probability under the null hypothesis and a test is of *level* α if its size is less than or equal to α.

[6] Note that the definition of the parameter space Γ is flexible. In some cases, one might want to define Γ so as to bound γ away from points in R^p that are troublesome. This is reasonable, of course, only if one has prior information that justifies the particular definition of Γ.

To a lesser extent, we are also interested in the minimum rejection probability of the test and its limit:

$$MinRP_n(\theta_0) = \inf_{\gamma \in \Gamma} RP_n(\theta_0, \gamma) \text{ and } AsyMinRP(\theta_0) = \liminf_{n \to \infty} MinRP_n(\theta_0). \quad (3.12)$$

The quantity $\alpha - MinRP_n(\theta_0)$ is the maximum amount of under-rejection of the test over points in the null hypothesis for fixed n. If $\alpha - AsyMinRP(\theta_0) > 0$, then the subsample test is not asymptotically similar and, hence, may sacrifice power.

4 Simple Model and Results

In this section, we use a simple model to illustrate the ideas behind the general results that are given below. We are interested in testing $H_0 : \theta = \theta_0$ in the presence of a nuisance parameter γ. The main simplifying assumptions made in this section are that γ is a scalar, Assumption Sub2 holds, and all asymptotic distributions are continuous and strictly increasing at their $1 - \alpha$ quantiles. These assumptions are too restrictive for most examples of practical interest, but imposing them is useful to illustrate the ideas. We suppose that the parameter space for γ is one-sided, i.e., $\Gamma = (0, b]$ or $[0, b]$ for some $0 < b < \infty$. The asymptotic distribution of the test statistic, $T_n(\theta_0)$, is assumed to be discontinuous at $\gamma = 0$ as defined precisely in Assumption B2 below.

Let $r > 0$ denote a *rate of convergence index.* When the sequence of true parameter values $\{\gamma_n : n \geq 1\}$ satisfies $n^r \gamma_n \to h \in [0, \infty]$, then the test statistic $T_n(\theta_0)$ is assumed to have an asymptotic distribution that depends on the localization parameter h, see Assumption B2 below. In most examples with i.i.d. observations, $r = 1/2$. The parameter space for the localization parameter h is $H = [0, \infty]$. Given $r > 0$ and $h \in H$, let $\{\gamma_{n,h} \in \Gamma : n \geq 1\}$ denote a sequence that satisfies $n^r \gamma_{n,h} \to h$. The upper curve in Figure 3(a) illustrates such a sequence for the case where $h < \infty$—the sequence decreases at rate n^{-r} for $r = 1/2$. The upper curve in Figure 3(b) illustrates such a sequence for the case where $h = \infty$—the sequence decreases at a rate slower than n^{-r} for $r = 1/2$. Note that a sequence $\{\gamma_{n,h} : n \geq 1\}$ that is bounded away from zero also satisfies $n^r \gamma_{n,h} \to h$ with $h = \infty$.

We make the following assumption.

Assumption B2. 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 $\{\gamma_{n,h} : n \geq 1\}$.

This assumption only requires verification of *pointwise* convergence of $T_n(\theta_0)$—no uniformity over γ parameters is required. Hence, verification can be carried out by the usual methods for determining the asymptotic distribution of a test statistic. Verification of Assumption B2 requires that one determines the asymptotic distributions of $T_n(\theta_0)$ under various sequences of true parameters, rather than under a fixed true parameter. But, this is done routinely for local power calculations. Assumption B2 requires that one considers all possible convergent sequences (possibly to ∞) of parameters in Γ, rather than just sequences of the form $\gamma_n = h/n^r$ for some h.

For a fixed critical value test, no further assumptions are needed to determine $AsySz(\theta_0)$. For subsample tests, we use the following additional assumption.

Assumption S. (i) $b_n \to \infty$, (ii) $b_n/n \to 0$, (iii) $\{T_{n,b_n,j}(\theta_0) : j = 1, ..., q_n\}$ and $T_{b_n}(\theta_0)$ are identically distributed under any $\gamma \in \Gamma$ for all $n \geq 1$, (iv) for all sequences $\{\gamma_n \in \Gamma : n \geq 1\}$, $U_{n,b_n}(x) - E_{\theta_0,\gamma_n} U_{n,b_n}(x) \to_p 0$ under $\{\gamma_n : n \geq 1\}$ for all $x \in R$, (v) for all $h \in H$, $J_h(x)$ is continuous and strictly increasing at its $1-\alpha$ quantile $c_h(1-\alpha)$, and (vi) Assumption Sub2 holds.

Assumptions S(i)-S(iii) are standard assumptions in the subsample literature. Assumption S(iv) automatically holds for i.i.d. observations (by a U-statistic inequality of Hoeffding as in PRW, p. 44). It also holds under stationary strong mixing observations given conditions on the mixing numbers, see below. Assumptions S(v) and S(vi) are imposed in this section for simplicity. They can be restrictive, so they are relaxed in the general results given below.

The asymptotic size of a subsample test is determined by the asymptotic distributions of the full-sample statistic $T_n(\theta_0)$ and the subsample statistic $T_{n,b_n,j}(\theta_0)$ under sequences $\{\gamma_{n,h} : n \geq 1\}$. By Assumption B2, the asymptotic distribution of $T_n(\theta_0)$ is J_h. The asymptotic distribution of $T_{n,b_n,j}(\theta_0)$ under $\{\gamma_{n,h} : n \geq 1\}$ is shown below to be J_g for some $g \in H$. Under $\{\gamma_{n,h} : n \geq 1\}$ for $h \in H$, not all $g \in H$ are possible indices for the asymptotic distribution of $T_{n,b_n,j}(\theta_0)$. The set of all possible pairs of localization parameters (g,h) is denoted GH and is defined by

$$GH = \{(g,h) \in H \times H : g = 0 \text{ if } h < \infty \ \& \ g \in [0,\infty] \text{ if } h = \infty\}. \tag{4.1}$$

Note that $g \leq h$ for all $(g,h) \in GH$.

Figure 3 provides an explanation for the form of GH and for the asymptotic behavior of the subsample test statistics. The upper curve in Figure 3(a) is the graph of $(n, \gamma_{n,h})$ for $n > 0$ when $\gamma_{n,h} = h/n^r$ for $0 < h < \infty$ and $r = 1/2$. The asymptotic distribution of $T_n(\theta_0)$ under this sequence is J_h. The question is "What is the asymptotic distribution of $T_{n,b_n,j}(\theta_0)$ under $\{(\gamma_{n,h} : n \geq 1\}$?" Because $T_{n,b_n,j}(\theta_0)$ has the same distribution as $T_{b_n}(\theta_0)$ by Assumption S(iii), the answer can be determined using Assumption B2 by considering the relationship between $\gamma_{n,h}$ and b_n.

For specificity, suppose $b_n = n^{1/2}$. (In practice b_n must be an integer, but for simplicity we ignore this in the present discussion.) The lower curve in Figure 3(a) is the graph of the pairs $(b_n, \gamma_{n,h}) = (n^{1/2}, h/n^{1/2})$ for $n > 0$ and $h = 1$ or, equivalently, the pairs $(n, h/n)$ for $n > 0$. The asymptotic distribution of $T_{b_n}(\theta_0)$ under $\{h/n^{1/2} : n \geq 1\}$ is the same as the asymptotic distribution of $T_n(\theta_0)$ under $\{h/n : n \geq 1\}$ because both correspond to the lower curve in Figure 3(a). The asymptotic distribution is J_0 by Assumption B2 because $n^r(h/n) = h/n^{1/2} \to 0$. Hence, if $T_n(\theta_0) \to_d J_h$ for $0 < h < \infty$, then $T_{n,b_n,j}(\theta_0) \to_d J_g$ for $g = 0$. This explains the condition in GH that $g = 0$ if $0 < h < \infty$.

Figure 3(a) illustrates the case in which $(g,h) = (0,h)$ for $0 < h < \infty$. Figure 3(b) illustrates an important second case in which $(g,h) = (g,\infty)$ for $0 < g < \infty$.

In Figure 3(b) the upper curve is the graph of $(n, \gamma_{n,h}) = (n, g/n^{1/4})$ for $n > 0$ and $g = 1$, and the lower curve is the graph of the pairs $(b_n, \gamma_{n,h}) = (n^{1/2}, g/n^{1/4})$ for $n > 0$ or, equivalently, the pairs $(n, g/n^{1/2})$ for $n > 0$. (Note that distributions under the parameters $\{\gamma_{n,h} = g/n^{1/4} : n \geq 1\}$ typically are not contiguous to those under $\gamma = 0$.) The lower curve in Figure 3(b) is the same as the upper curve in Figure 3(a). The asymptotic distribution of $T_{b_n}(\theta_0)$ under $\{g/n^{1/4} : n \geq 1\}$ is the same as the asymptotic distribution of $T_n(\theta_0)$ under $\{g/n^{1/2} : n \geq 1\}$ because both correspond to the lower curve in Figure 3(b). This asymptotic distribution is J_g by Assumption B2 because $n^r(g/n^{1/2}) = g \to g$. In consequence, in this case, $T_n(\theta_0) \to_d J_h$ for $h = \infty$ and $T_{n,b_n,j}(\theta_0) \to_d J_g$ for $0 < g < \infty$. Thus, if $h = \infty$, then g can take any value in $(0, \infty)$, as is allowed in GH. (By considering the sequences $\gamma_{n,h} = 1/n$, $1/n^{1/3}$, and $1/n^{1/5}$, one can show that the pairs $(g, h) = (0, 0)$, $(0, \infty)$ and (∞, ∞) also are possible and, hence, are contained in GH.)

In sum, under a sequence $\{\gamma_{n,h} : n \geq 1\}$, the asymptotic distributions of $T_n(\theta_0)$ and $T_{n,b_n,j}(\theta_0)$ are given by J_h and J_g, respectively, where $g = \lim_{n\to\infty} b_n^r \gamma_{n,h}$. The possible (g, h) pairs are those contained in the localization parameter space GH.

Using the heuristics given above, Corollary 1 below shows that the asymptotic sizes of FCV and subsample tests are given by the following quantities:

$$\begin{aligned} Max_{Fix}(\alpha) &= \sup_{h \in H} [1 - J_h(c_{Fix}(1-\alpha))] \text{ and} \\ Max_{Sub}(\alpha) &= \sup_{(g,h) \in GH} [1 - J_h(c_g(1-\alpha))], \end{aligned} \quad (4.2)$$

respectively, where $c_g(1-\alpha)$ denotes the $1-\alpha$ quantile of J_g. The Corollary applies to the simple model considered in this section. It is a special case of Theorem 2 below. For the reader's convenience, a proof of Corollary 1 is given in Appendix B. This proof is simpler than that of Theorem 2 but contains the main ideas of the latter.

Corollary 1 *Suppose Assumption B2 holds and* $\Gamma = (0, b]$ *or* $[0, b]$ *for some* $0 < b < \infty$. *Then,*

(a) $AsySz(\theta_0) = Max_{Fix}(\alpha)$ *for an FCV test, and*

(b) $AsySz(\theta_0) = Max_{Sub}(\alpha)$ *for a subsample test provided Assumption S holds.*

Comments. 1. A key question concerning nominal level α FCV and subsample tests is whether $AsySz(\theta_0) \leq \alpha$. For an FCV test, Corollary 1 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, denoted $c_h(1-\alpha)$, for all $h \in H$.

2. For a subsample test, Corollary 1 shows that $AsySz(\theta_0) \leq \alpha$ iff $c_g(1-\alpha) \geq c_h(1-\alpha)$ for all $(g, h) \in GH$ (because in the latter case $1 - J_h(c_g(1-\alpha)) \leq 1 - J_h(c_h(1-\alpha)) = \alpha$). In consequence, a graph of $c_h(1-\alpha)$ as a function of h is very informative concerning the asymptotic size of a subsample test. Figure 4 provides four examples of possible shapes of $c_h(1-\alpha)$ as a function of h (each of which arises in real applications of interest). In Figure 4(a), $c_h(1-\alpha)$ is strictly decreasing in h. Hence, $c_g(1-\alpha) \geq c_h(1-\alpha)$ for all $(g,h) \in GH$ (since $g \leq h$) and $AsySz(\theta_0) \leq \alpha$. In contrast, in Figures 4(b), 4(c), and 4(d), there are pairs $(g,h) \in GH$ for which $c_g(1-\alpha) < c_h(1-\alpha)$ and, hence, $AsySz(\theta_0) > \alpha$. In Figure 4(b), this is true for all (g,h) with $g = 0$ and $0 < h < \infty$ or $0 \leq g < \infty$ and $h = \infty$ because $c_h(1-\alpha)$ is strictly increasing in h. In Figure 4(c), it is true for all (g,h) with $0 < g < \infty$ and $h = \infty$. In Figure 4(d), it is true for all (g,h) with $g = 0$ and $0 < h < \infty$. Figure 4(c) illustrates a case in which subsampling does not lead to over-rejection for alternatives that typically are contiguous, i.e., those with $h < \infty$, but leads to over-rejection for alternatives that typically are not contiguous, i.e., $h = \infty$. In general, by looking at a graph of $c_h(1-\alpha)$ as a function of h, one can see which values of h lead to over- or under-rejection of the null. The shape of the graph determines whether $AsySz(\theta_0) \leq \alpha$ for a subsample test.

3. Typically $AsySz(\theta_0)$ for an FCV test is relatively easy to compute by simulation of $J_h(c_{Fix}(1-\alpha))$ for all $h \in H$. Similarly, for subsample tests, usually one can calculate $AsySz(\theta_0)$ via simulation without great difficulty.

5 Assumptions

The previous section provides assumptions and results for a simple model for illustrative purposes. In this section, we introduce the general assumptions that are employed in the paper. The assumptions are verified in several examples below.

5.1 Parameter Space

First, we introduce some notation. 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. Let $R_+ = \{x \in R : x \geq 0\}$, $R_- = \{x \in R : x \leq 0\}$, $R_{+,\infty} = R_+ \cup \{\infty\}$, $R_{-,\infty} = R_- \cup \{-\infty\}$, $R_\infty = R \cup \{\pm\infty\}$, $R_+^p = R_+ \times ... \times R_+$ (with p copies), and $R_\infty^p = R_\infty \times ... \times R_\infty$ (with p copies).

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, $\gamma_1 \in R^p$. Through reparametrization we can assume without loss of generality that the discontinuity occurs when one or more elements of γ_1 equal zero. The value of γ_1 affects the limit distribution of the test statistic of interest. The parameter space for γ_1 is $\Gamma_1 \subset R^p$.

The second component, γ_2 ($\in R^q$), of γ 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 $\Gamma_2 \subset R^q$.[7]

The third component, γ_3, of γ does not affect the limit distribution of the test statistic. It is assumed to be an element of an arbitrary space $\mathcal{T}_3$. Hence, it may be finite

[7] The extension to the case where γ_2 is infinite dimensional is straightforward. In the examples we consider, γ_2 is finite dimensional. So, for simplicity, we take it to be so here.

or infinite dimensional. For example, in a linear model, a test statistic concerning one regression parameter may be invariant to the value of some other regression parameters. The latter parameters are then part of γ_3. Infinite dimensional γ_3 parameters also arise frequently. For example, error distributions are often part of γ_3. Due to the operation of the central limit theorem (CLT) it is often the case that the asymptotic distribution of a test statistic does not depend on the particular error distribution—only on whether the error distribution has certain moments finite. Such error distributions are part of γ_3. The parameter space for γ_3 is $\Gamma_3(\gamma_1, \gamma_2)$ ($\subset \mathcal{T}_3$), which as indicated may depend on γ_1 and γ_2.

The parameter space for γ is

$$\Gamma = \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\}. \tag{5.1}$$

In Section 6 below we provide two main theorems. The first theorem relies on weaker assumptions than the second, but gives weaker results. An assumption label that ends with 1 is used in Theorem 1. An assumption label that ends in 2 is used in Theorem 2 and is stronger than a corresponding assumption that ends in 1. (For example, Assumption A2 implies Assumption A1.) All other assumptions are used in both Theorems 1 and 2.

Assumption A1. Γ satisfies (5.1), where $\Gamma_1 \subset R^p$, $\Gamma_2 \subset R^q$, and $\Gamma_3(\gamma_1, \gamma_2) \subset \mathcal{T}_3$ for some arbitrary space $\mathcal{T}_3$.

Assumption A2. (i) Assumption A1 holds and (ii) $\Gamma_1 = \prod_{m=1}^{p} \Gamma_{1,m}$, where $\Gamma_{1,m} = \lfloor a_m, b_m \rfloor$ for some $-\infty \leq a_m < b_m \leq \infty$ that satisfy $a_m \leq 0 \leq b_m$ for $m = 1, ..., p$.

Under Assumption A2, the parameter space Γ_1 includes values γ_1 that are arbitrarily close to 0.[8]

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 A2, set $\Gamma_1 = R_+$, $\Gamma_2 = (-1, 1)$, and $\Gamma_3(\gamma_1, \gamma_2) = (0, \infty) \times (0, \infty) \times \mathcal{F}(\mu, \rho, \sigma_1, \sigma_2)$, where

$$\begin{aligned}\mathcal{F}(\mu, \rho, \sigma_1, \sigma_2) = \{F : E_F||X_i||^{2+\delta} \leq M,\ E_F X_i = (0, \mu)',\ Var_F(X_{i1}) = \sigma_1^2, \\ Var_F(X_{i2}) = \sigma_2^2,\ \&\ Corr_F(X_{i1}, X_{i2}) = \rho\}\end{aligned} \tag{5.2}$$

for some $M < \infty$ and $\delta > 0$.[9] Then, Assumption A2 holds.

The null distribution of $T_n^*(\theta_0)$ is invariant to σ_1^2 because $\widehat{\sigma}_{n1}$ is scale equivariant. Hence, 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 $\Gamma_3(\gamma_1, \gamma_2) = \{1\} \times (0, \infty) \times \mathcal{F}(\mu, \rho, \sigma_1, \sigma_2)$.

[8] The results below allow for the case where there is no subvector γ_1 of γ, i.e., $p = 0$. In this case, there is no discontinuity of the asymptotic distribution of the test statistic of interest, see below.

[9] The condition $E_F||X_i||^{2+\delta} \leq M$ in $\mathcal{F}(\mu, \rho, \sigma_1, \sigma_2)$ ensures that the Liapunov CLT applies in (5.5)-(5.7) below. In $\mathcal{F}(\mu, \rho, \sigma_1, \sigma_2)$, $E_F X_{i1} = 0$ because the results given are all under the null hypothesis.

5.2 Convergence Assumptions

This subsection and the next introduce the high-level assumptions that we employ. The high-level assumptions are verified in several examples below.

Throughout this section, the true value of θ is the null value θ_0 and all limits are as $n \to \infty$. For an arbitrary distribution G, let $G(\cdot)$ denote the distribution function (df) of G and let $C(G)$ denote the continuity points of $G(\cdot)$. Define the $1-\alpha$ quantile, $q(1-\alpha)$, of a distribution G by $q(1-\alpha) = \inf\{x \in R : G(x) \geq 1-\alpha\}$. For a df $G(\cdot)$, let $G(x-) = \lim_{\varepsilon \searrow 0} G(x-\varepsilon)$, where "$\lim_{\varepsilon \searrow 0}$" denotes the limit as $\varepsilon > 0$ declines to zero. Note that $G(x+) = \lim_{\varepsilon \searrow 0} G(x+\varepsilon)$ equals $G(x)$ because dfs are right continuous. The distributions J_h and J_{h^0} considered below are distributions of proper random variables that are finite with probability one.

For a sequence of constants $\{\kappa_n : n \geq 1\}$, let $\kappa_n \to [\kappa_{1,\infty}, \kappa_{2,\infty}]$ denote that $\kappa_{1,\infty} \leq \liminf_{n\to\infty} \kappa_n \leq \limsup_{n\to\infty} \kappa_n \leq \kappa_{2,\infty}$.

Let $r > 0$ denote a *rate of convergence index* such that when the true value of 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. In most examples, $r = 1/2$, but in the unit root example considered in Andrews and Guggenberger (2005a) $r = 1$.

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

$$\begin{aligned} H = \{&h = (h_1, h_2) \in R_\infty^{p+q} : \exists \, \{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\} \\ &\text{such that } n^r \gamma_{n,1} \to h_1 \text{ and } \gamma_{n,2} \to h_2\}. \end{aligned} \tag{5.3}$$

For notational simplicity, in the definition of H and below, we write (h_1, h_2), rather than $(h_1', h_2')'$, even though h is a $p+q$ column vector. Under Assumption A2, it follows that

$$H = H_1 \times H_2, \; H_1 = \prod_{m=1}^{p} \begin{cases} R_{+,\infty} & \text{if } a_m = 0 \\ R_{-,\infty} & \text{if } b_m = 0 \\ R_\infty & \text{if } a_m < 0 \text{ and } b_m > 0, \end{cases} \quad \text{and } H_2 = \text{cl}(\Gamma_2), \tag{5.4}$$

where $\text{cl}(\Gamma_2)$ is the closure of Γ_2 with respect to R_∞^q. For example, if $p = 1$, $a_1 = 0$, 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 $\gamma_{n,h,2} \to h_2$.

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, see Assumptions B1 and B2 below. For a given model, there is a single fixed $r > 0$. Hence, for notational simplicity, we do not index $\{\gamma_{n,h} : n \geq 1\}$ by r. In addition, the limit distributions under $\{\gamma_{n,h} : n \geq 1\}$ of the test statistics of interest do not depend on $\gamma_{n,h,3}$, so we do not make the dependence of $\gamma_{n,h}$ on $\gamma_{n,h,3}$ explicit.

In models in which the asymptotic distribution of the test statistic of interest is continuous in the model parameters, we apply the results below with no parameter γ_1 (or $\gamma_{n,h,1}$), i.e., $p = 0$. We refer to this as the *continuous limit* case. In the continuous limit case, the asymptotic distribution of the test statistic is the same for all sequences of true parameters that converge to the same point. On the other hand, in the *discontinuous limit* case, which is the case of main interest in this paper, we apply the results with $p \geq 1$.

Given any $h = (h_1, h_2) \in H$, define $h^0 = (0, h_2)$.

We use the following assumptions.

Assumption B1. (i) For some $r > 0$, some $h \in H \cap R^{p+q}$ such that $h^0 \in H$, some sequence $\{\gamma_{n,h} : n \geq 1\}$, and some distribution J_h, $T_n(\theta_0) \to_d J_h$ under $\{\gamma_{n,h} : n \geq 1\}$ and (ii) for all sequences $\{\gamma_{n,h^0} : n \geq 1\}$ and some distribution J_{h^0}, $T_n(\theta_0) \to_d J_{h^0}$ under $\{\gamma_{n,h^0} : n \geq 1\}$.

Assumption B2. 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 $\{\gamma_{n,h} : n \geq 1\}$.

If $\gamma_{n,h}$ does not depend on n (which necessarily requires $h_1 = 0$), Assumption B1(i) is a standard assumption in the subsample literature. For example, it is imposed in the basic theorem in PRW, Thm. 2.2.1, p. 43, for subsampling with i.i.d. observations and in their Thm. 3.2.1, p. 70, for stationary strong mixing observations. If $\gamma_{n,h}$ does depend on n, Assumption B1(i) usually can be verified using the same sort of argument as when it does not. Similarly, Assumption B1(ii) usually can be verified using the same sort of argument and, hence, is not restrictive.

Assumption B2 holds in many examples, but it can be restrictive. It is for this reason that we introduce Assumption B1. Theorem 1 only requires Assumption B1, whereas Theorem 2 requires Assumption B2. In the "continuous limit" case (where Assumption B2 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 B2 for this example. For more complicated boundary examples, results of Andrews (1999, 2001) can be used to verify Assumption B2. 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

$$\begin{pmatrix} n^{1/2}\overline{X}_{n1}/\widehat{\sigma}_{n1} \\ n^{1/2}\overline{X}_{n2}/\widehat{\sigma}_{n2} \end{pmatrix} \to_d \begin{pmatrix} 0 \\ h_1 \end{pmatrix} + Z_{h_2}, \tag{5.5}$$

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 $\{\gamma_{n,h} = (\mu_{n,h}/\sigma_{2,n,h}, \rho_{n,h}, (\sigma_{1,n,h}, \sigma_{2,n,h}, F_{n,h})) : n \geq 1\}$.)

By the continuous mapping theorem, we obtain

$$T_n^*(\theta_0) = n^{1/2}\widehat{\theta}_n/\widehat{\sigma}_{n1} = n^{1/2}\overline{X}_{n1}/\widehat{\sigma}_{n1} - \widehat{\rho}_n \min(0, n^{1/2}\overline{X}_{n2}/\widehat{\sigma}_{n2})) \to_d J_h^* \tag{5.6}$$

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

$$Z_{h_2,1} - h_2 \min(0, Z_{h_2,2} + h_1). \tag{5.7}$$

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 $h_2 < 0$ ($h_2 \geq 0$).

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

$$T_n^*(\theta_0) \to_d J_\infty^*, \text{ where } J_\infty^* \text{ is the } N(0,1) \text{ distribution.} \tag{5.8}$$

Note that the limit distributions J_h^* and J_∞^* do not depend on $\gamma_3 = (\sigma_1^2, \sigma_2^2, F)$.

For $T_n(\theta_0) = T_n^*(\theta_0), -T_n^*(\theta_0)$, and $|T_n^*(\theta_0)|$, we have $T_n(\theta_0) \to J_h$ under $\{\gamma_{n,h}\}$, where $J_h = J_h^*, -J_h^*$, and $|J_h^*|$, respectively. (If $Y \sim J_h^*$, then by definition, $-Y \sim -J_h^*$ and $|Y| \sim |J_h^*|$.) Hence, Assumption B2 holds for upper, lower, and symmetric tests.

Figure 5 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 2 to Corollary 1, these graphs provide considerable qualitative information concerning the null rejection probabilities of subsample 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 ($= \lim_{n\to\infty} \rho_{n,h}$).

The quantile graphs for J_h^* indicate that the upper one-sided subsample 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 subsample test under-rejects (because the graphs are decreasing in h_1) with the greatest degree of under-rejection 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 subsample test exceeds .05 by a substantial amount.

In contrast, the graphs for J_h^* compared with the standard fixed critical value of 1.64 show that the upper one-sided FCV test under-rejects for negative values of h_2 for all values of h_1 (because the graphs lie below 1.64) with the greatest degree of under-rejection being quite large and occurring for h_2 close to -1. Note that the maximum null rejection probability over $h_1 \in [0, \infty)$ for any given $h_2 \leq 0$ is .05 because the quantile graphs are maximized at $h_1 = \infty$ and asymptote to 1.64 as $h_1 \to \infty$. For positive values of h_2, the upper FCV test over-rejects (because the graphs lie above 1.64) with the greatest degree of over-rejection being relatively small and occurring for $h_1 = 0$ and h_2 around .5. Hence, the graphs indicate that the size of the upper FCV test is greater than .05, but by a modest amount. (The results for lower one-sided subsample and FCV tests are analogous to those for upper one-sided tests with h_2 replaced by $-h_2$.)

The quantile graphs for $|J_h^*|$ indicate that the symmetric two-sided subsample test over-rejects for non-zero values of h_2 for all values of h_1 (because the graphs are increasing in h_1) with the greatest degree of over-rejection being moderate and occurring for $(g_1, h_1) = (0, \infty)$ and $|h_2|$ close to 1. (Note that the graphs for $|J_h^*|$ are invariant to the sign of h_2.) The graphs indicate that the size of the symmetric subsample test exceeds .05 by a moderate amount.

The graphs for $|J_h^*|$ compared with the standard fixed critical value of 1.96 show that the symmetric two-sided FCV test under-rejects for non-zero values of h_2 for all finite values of h_1 (because the graphs lie below 1.96) with the greatest degree of under-rejection being modest and occurring for $h_1 = 0$ and $|h_2|$ close to 1. The maximum null rejection probability over $h_1 \in [0, \infty)$ for any given $h_2 \leq 0$ is .05 (for the same reason as given above for J_h). The graphs indicate that the asymptotic size of the symmetric FCV test is .05, as desired, but the test is not asymptotically similar.

5.3 Subsample Assumptions

The assumptions above are all that are needed for FCV tests. For subsample tests, we require the following additional assumptions:

Assumption C. (i) $b_n \to \infty$ and (ii) $b_n/n \to 0$.

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

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

Assumption F1. For all $\varepsilon > 0$, $J_{h^0}(c_{h^0}(1-\alpha) + \varepsilon) > 1 - \alpha$, where $c_{h^0}(1-\alpha)$ is the $1 - \alpha$ quantile of J_{h^0} and h^0 is as in Assumption B1(ii).

Assumption F2. 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 G1. For the sequence $\{\gamma_{n,h} : n \geq 1\}$ in Assumption B1(i), $L_{n,b_n}(x) - U_{n,b_n}(x) \to_p 0$ for all $x \in C(J_{h^0})$ under $\{\gamma_{n,h} : n \geq 1\}$.

Assumption G2. For all $h = (h_1, h_2) \in H$ and all sequences $\{\gamma_{n,h} : n \geq 1\}$ for which $b_n^r \gamma_{n,h,1} \to g_1$ for some $g_1 \in R_\infty^p$, if $U_{n,b_n}(x) \to_p J_g(x)$ 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 $L_{n,b_n}(x) - U_{n,b_n}(x) \to_p 0$ under $\{\gamma_{n,h} : n \geq 1\}$ for all $x \in C(J_g)$.

Assumptions C and D are standard assumptions in the subsample literature, e.g., see PRW, Thm. 2.2.1, p. 43, and are not restrictive. The sequence $\{b_n : n \geq 1\}$ can always be chosen to satisfy Assumption C. Assumption D necessarily holds when the observations are i.i.d. or stationary and subsamples are constructed in the usual way (described above).

Assumption E holds quite generally. For i.i.d. observations, the condition in Assumption E when $\gamma_{n,1} = 0$ and $(\gamma_{n,2}, \gamma_{n,3})$ does not depend on n (where $\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3})$) is verified by PRW, p. 44, using a U-statistic inequality of Hoeffding. It 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 $\gamma \in \Gamma$ when the subsample statistics are defined as above.

For stationary strong mixing observations, the condition in Assumption E when $\gamma_{n,1} = 0$ and $(\gamma_{n,2}, \gamma_{n,3})$ does not depend on n (where $\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3})$) is verified by 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

$$\sup_{\gamma \in \Gamma} \alpha_\gamma(m) \to 0 \text{ as } m \to \infty, \tag{5.9}$$

where $\{\alpha_\gamma(m) : m \geq 1\}$ are the strong mixing numbers of the observations when the true parameters are (θ_0, γ).

Assumptions F1 and F2 are designed to avoid the requirement that $J_h(x)$ is continuous in x because this assumption is violated in some applications, such as with super-efficient estimators and some boundary problems, for some values of h and some values of x. Assumptions F1 and F2 hold in all of the examples that we have considered. In particular, Assumption F1 holds if either (i) $J_{h^0}(x)$ is strictly increasing at $x = c_{h^0}(1-\alpha)$ or (ii) $J_{h^0}(x)$ has a jump at $x = c_{h^0}(1-\alpha)$ with $J_{h^0}(c_{h^0}(1-\alpha)) > 1-\alpha$. Condition (i) holds in most examples. But, if J_{h^0} is a pointmass, as occurs with the example of a CI based on a super-efficient estimator with constant $a = 0$ (see Section 9.1 below), then condition (i) fails, but condition (ii) holds. Sufficient conditions for Assumption F2 are the same with h^0 replaced by h for all $h \in H$.

Assumptions G1 and G2 hold automatically when $\{\widehat{T}_{n,b_n,j}\}$ satisfy Assumption Sub2. To verify that Assumption G1 or G2 holds when $\{\widehat{T}_{n,b_n,j}\}$ satisfy Assumption Sub1 and $T_n(\theta_0)$ is a non-studentized t statistic (i.e., Assumptions t1 and t2 hold), we use the following assumption.

Assumption H. $\tau_{b_n}/\tau_n \to 0$.

This is a standard assumption in the subsample literature, e.g., see PRW, Thm. 2.2.1, p. 43. In the leading case where $\tau_n = n^s$ for some $s > 0$, Assumption H follows from Assumption C(ii) because $\tau_{b_n}/\tau_n = (b_n/n)^s \to 0$.

Lemma 1 (a) *Assumptions* B1(i), t1, t2, Sub1, *and* H *imply Assumption* G1.
(b) *Assumptions* B2, t1, t2, Sub1, *and* H *imply Assumption* G2.

Comment. Lemma 1 is a special case of Lemma 2, which is stated in Appendix A for expositional convenience. Lemma 2 does not impose Assumption t2 and, hence, covers studentized t statistics.

Example 1 (cont.). We now verify Assumptions C-F2 for this example. Assumptions C and D clearly hold. Assumption E holds by the general argument given above. For $\alpha < 1/2$, Assumption F2 holds for $J_h = J_h^*$ (defined above in (5.7)-(5.8)) 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 F2 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 $Z \sim N(0,1)$). Assumption G2 is verified in Appendix A following the statement of Lemma 2.

6 Asymptotic Results

The first result of this section concerns the asymptotic null behavior of FCV and subsample tests under a single sequence $\{\gamma_{n,h} : n \geq 1\}$.

Theorem 1 (a) *Suppose Assumption* B1(i) *holds. Then,*

$$P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{Fix}(1-\alpha)) \to [1 - J_h(c_{Fix}(1-\alpha)), 1 - J_h(c_{Fix}(1-\alpha)-)].$$

(b) *Suppose Assumptions* A1, B1, C-E, F1, *and* G1 *hold. Then,*

$$P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{n,b}(1-\alpha)) \to [1 - J_h(c_{h^0}(1-\alpha)), 1 - J_h(c_{h^0}(1-\alpha)-)].$$

Comments. 1. If $1 - J_h(c_{Fix}(1-\alpha)) > \alpha$, then the FCV test has $AsySz(\theta_0) > \alpha$, i.e., its asymptotic size exceeds its nominal level α.

2. Analogously, for the subsample test, if $1 - J_h(c_{h^0}(1-\alpha)) > \alpha$, then the test has $AsySz(\theta_0) > \alpha$.

3. If $1 - J_h(c_{Fix}(1-\alpha)-) < \alpha$, then the FCV test has $AsyMinRP(\theta_0) < \alpha$ and it is not asymptotically similar. Analogously, if $1 - J_h(c_{h^0}(1-\alpha)-) < \alpha$, then the subsample test has $AsyMinRP(\theta_0) < \alpha$ and it is not asymptotically similar.

4. If $J_h(x)$ is continuous at $x = c_{Fix}(1-\alpha)$ (which typically holds in applications for most values h but not necessarily all), then the result of Theorem 1(a) becomes $P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{Fix}(1-\alpha)) \to 1 - J_h(c_{Fix}(1-\alpha))$. Analogously, if $J_h(x)$ is continuous at $x = c_{h^0}(1-\alpha)$, then the result of Theorem 1(b) becomes $P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{n,b}(1-\alpha)) \to 1 - J_h(c_{h^0}(1-\alpha))$.

5. In the "continuous limit" case, $h^0 = h$ because no parameter γ_1 appears. Hence, the result of Theorem 1(b) for the subsample test is $P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{n,b}(1-\alpha)) \to \alpha$, provided $J_h(x)$ is continuous at $x = c_h(1-\alpha)$. That is, the pointwise asymptotic null rejection rate is the desired nominal rate α.

6. Typically Assumption B1(i) holds for an infinite number of values h, say $h \in H^*$ ($\subset R^p$). In this case, Comments 1-5 apply for all $h \in H^*$.

We now use the stronger Assumptions A2, B2, F2, and G2 to establish more precisely the asymptotic sizes and asymptotic minimum rejection probabilities of sequences of FCV and subsample tests. For FCV tests, we define

$$
\begin{aligned}
Max_{Fix}(\alpha) &= \sup_{h \in H} [1 - J_h(c_{Fix}(1-\alpha))] \text{ and} \\
Max_{Fix}^{-}(\alpha) &= \sup_{h \in H} [1 - J_h(c_{Fix}(1-\alpha)-)].
\end{aligned} \tag{6.1}
$$

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

For subsample tests, define

$$
\begin{aligned}
GH = \{&(g,h) \in H \times H : g = (g_1, g_2),\ h = (h_1, h_2),\ g_2 = h_2 \text{ and for } m = 1, ..., p, \\
&\text{(i) } g_{1,m} = 0 \text{ if } |h_{1,m}| < \infty,\ \text{(ii) } g_{1,m} \in R_{+,\infty} \text{ if } h_{1,m} = +\infty,\ \text{ and} \\
&\text{(iii) } g_{1,m} \in R_{-,\infty} \text{ if } h_{1,m} = -\infty\},
\end{aligned} \tag{6.2}
$$

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

Define

$$
\begin{aligned}
Max_{Sub}(\alpha) &= \sup_{(g,h) \in GH} [1 - J_h(c_g(1-\alpha))] \text{ and} \\
Max_{Sub}^{-}(\alpha) &= \sup_{(g,h) \in GH} [1 - J_h(c_g(1-\alpha)-)].
\end{aligned} \tag{6.3}
$$

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 $c_h(1-\alpha)$.

Theorem 2 (a) *Suppose Assumptions* A2 *and* B2 *hold. Then, an FCV test satisfies*

$$
\begin{aligned}
&AsySz(\theta_0) \in [Max_{Fix}(\alpha), Max_{Fix}^{-}(\alpha)] \text{ and} \\
&AsyMinRP(\theta_0) \in [Min_{Fix}(\alpha), Min_{Fix}^{-}(\alpha)].
\end{aligned}
$$

(b) *Suppose Assumptions* A2, B2, C-E, F2, *and* G2 *hold. Then, a subsample test satisfies*

$$
\begin{aligned}
&AsySz(\theta_0) \in [Max_{Sub}(\alpha), Max_{Sub}^{-}(\alpha)] \text{ and} \\
&AsyMinRP(\theta_0) \in [Min_{Sub}(\alpha), Min_{Sub}^{-}(\alpha)].
\end{aligned}
$$

Comments. 1. If $J_h(x)$ is continuous at the appropriate value(s) of x, then $Max_{Fix}(\alpha) = Max_{Fix}^{-}(\alpha)$ and $Max_{Sub}(\alpha) = Max_{Sub}^{-}(\alpha)$ and Theorem 2 gives the precise value of $AsySz(\theta_0)$ and analogously for $AsyMinRP(\theta_0)$. Even for FCV tests, we are not aware of general results in the literature that establish the limit of the finite-sample size of a test based on a test statistic whose limit distribution depends on nuisance parameters.

2. Given Theorem 2(b) and the definition of $Max^-_{Sub}(\alpha)$, sufficient conditions for a nominal level α subsample test to have asymptotic level α are the following: (a) $c_g(1-\alpha) \geq c_h(1-\alpha)$ for all $(g,h) \in GH$ and (b) $Max^-_{Sub}(\alpha) = Max_{Sub}(\alpha)$.[10] Condition (a) necessarily holds in "continuous limit" examples and it holds in some "discontinuous limit" examples. But, it often fails in "discontinuous limit" examples. Condition (b) holds in most examples.

3. The same argument as used to prove Theorem 2 can be used to prove slightly stronger results than those of Theorem 2. Namely, for FCV and subsample tests,

$$ExSz_n(\theta_0) \to [Max_{Type}(\alpha), Max^-_{Type}(\alpha)] \tag{6.4}$$

for $Type = Fix$ and Sub, respectively. (These results are stronger because they imply that $\liminf_{n\to\infty} ExSz_n(\theta_0) \geq Max_{Type}(\alpha)$, rather than just $AsySz(\theta_0) = \limsup_{n\to\infty} ExSz_n(\theta_0) \geq Max_{Type}(\alpha)$.) Hence, when $Max_{Type}(\alpha) = Max^-_{Type}(\alpha)$, we have

$$\lim_{n\to\infty} ExSz_n(\theta_0) = Max_{Type}(\alpha) \text{ for } Type = Fix \text{ and } Sub. \tag{6.5}$$

4. The upper bound on $AsySz(\theta_0)$ in Theorem 2(b) can be improved in some cases, see the fourth section of Appendix A.

Example 1 (cont.). The size properties of the tests in this example are given in Table I (which is described in more detail below) and are summarized as follows. For upper and lower one-sided tests, we find large asymptotic size distortions for the subsample 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 (lower) one-sided subsample test over-rejects the null most when the correlation h_2 ($= \lim_{n\to\infty} \rho_{n,h}$) is close to -1 (respectively, 1). Monte Carlo simulations of its asymptotic null rejection probabilities show that its asymptotic size is about $1/2$ for all nominal sizes $\alpha \in [.01, .2]$ that we consider.

The symmetric two-sided subsample test also is found to be size-distorted, but by a much smaller amount. The Monte Carlo simulations for $\alpha \in [.01, .2]$ show that $AsySz(\theta_0)$ is about 2α for the symmetric subsample test. 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. All of the quantitative results discussed above are consistent with the qualitative results obtained from the quantile graphs and discussed in Section 5.2.

The results described above are determined as follows. First, for upper one-sided, lower one-sided, and symmetric two-sided tests, $Max^-_{Type}(\alpha) = Max_{Type}(\alpha)$ for any $\alpha \in (0,1)$ for $Type = Fix$ and Sub. This is shown at the end of the first section of Appendix A.

[10] Under these conditions, $Max^-_{Sub}(\alpha) = Max_{Sub}(\alpha) = \sup_{(g,h)\in GH}[1 - J_h(c_g(1-\alpha))] \leq \sup_{h\in H}[1 - J_h(c_h(1-\alpha))] \leq \alpha$.

Second, we discuss upper one-sided tests. Given that $J_h = J_h^*$ is stochastically increasing (decreasing) in h_1 for fixed $h_2 < 0$ ($h_2 \geq 0$), one can show that

$$\begin{aligned} Max_{Fix}(\alpha) &= \sup_{h \in H}[1 - J_h(c_{Fix}(1-\alpha))] = \sup_{h_2 \in [0,1]}(1 - J_{(0,h_2)}(z_{1-\alpha})) \text{ and} \\ Max_{Sub}(\alpha) &= \sup_{(g,h) \in GH}[1 - J_h(c_g(1-\alpha))] = \sup_{h_2 \in [-1,0]}(1 - J_\infty(c_{(0,h_2)}(1-\alpha))), \end{aligned} \quad (6.6)$$

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. See Appendix A for proofs. 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.

The values of $Max_{Fix}(\alpha)$ and $Max_{Sub}(\alpha)$ for the upper and lower one-sided tests are obtained by simulation. (All simulation results are based on 50,000 simulation repetitions and when maximization over h_1 ($= \lim_{n\to\infty} n^{1/2}\mu_{n,h}/\sigma_{2,n,h}$) is needed the upper bound is 12 and a grid of size 0.05 is used.) Table I reports $1 - J_{(0,h_2)}(z_{1-\alpha})$ and $1 - J_\infty(c_{(0,h_2)}(1-\alpha))$ for FCV and subsample tests, respectively, for various values of h_2 ($= \lim_{n\to\infty} \rho_{n,h}$) and $\alpha = .05$ for upper one-sided tests. Because the results for lower one-sided tests are the same, but with h_2 replaced by $-h_2$, the results for lower one-sided tests are not reported in Table I. The last row of Table I gives the $AsySz(\theta_0)$ of each test, which is maximum of the numbers in each column. For one-sided tests, simulation of $AsySz(\theta_0)$ is very fast because the two-dimensional maximization over (h_1, h_2) has been reduced to a one-dimensional maximization over h_2 in (6.6).

Third, for the symmetric two-sided case, where $J_h = |J_h^*|$, we have

$$\begin{aligned} Max_{Fix}(\alpha) &= \sup_{h \in H}[1 - J_h(z_{1-\alpha/2})] \text{ and} \\ Max_{Sub}(\alpha) &= \max\{\sup_{h_1 \in [0,\infty), h_2 \in [-1,1]}[1 - J_h(c_{(0,h_2)}(1-\alpha))], \\ &\quad \sup_{h_1 \in [0,\infty], h_2 \in [-1,1]}[1 - J_\infty(c_{(h_1,h_2)}(1-\alpha))]\}. \end{aligned} \quad (6.7)$$

We use Monte Carlo simulation to calculate these quantities. For the FCV test, Table I reports the values of $\sup_{h_1 \in [0,\infty]}[1 - J_{(h_1,h_2)}(z_{1-\alpha/2})]$ appearing in $Max_{Fix}(\alpha)$ for a range of h_2 values in $[-1,1]$ and $\alpha = .05$. Note that these are the maximum asymptotic null rejection probabilities given h_2, where the maximum is over h_1 with h_2 fixed. Table I also reports the analogous expressions that depend on h_2 for the subsample tests. Table I shows that $AsySz(\theta_0)$ is about 2α for the symmetric two-sided subsample tests and α for the FCV tests.

In conclusion, the one-sided subsample test has very large size-distortion, whereas the one-sided FCV test has very small size distortion. The symmetric two-sided subsample test has moderate size distortion, whereas symmetric FCV has no size distortion.

7 Equal-tailed t Tests

This section considers *equal-tailed* two-sided t tests. There are two reasons for considering such tests. First, equal-tailed tests and CIs are preferred to symmetric procedures by some statisticians, e.g., see Efron and Tibshirani (1993). Second, given the potential problems of symmetric t tests documented in Section 6, it is of interest to see whether equal-tailed tests are subject to the same problems and, if so, whether the problems are more or less severe than for symmetric procedures.

We suppose Assumption t1(i) holds, so that $T_n(\theta_0) = \tau_n(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n$. An equal-tailed FCV or subsample t test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ of nominal level α ($\in (0, 1/2)$) rejects H_0 when

$$T_n(\theta_0) > c_{1-\alpha/2} \text{ or } T_n(\theta_0) < c_{\alpha/2}, \tag{7.1}$$

where $c_{1-\alpha}$ is defined in (3.10) for FCV and subsample tests.

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

$$ExSz_n(\theta_0) = \sup_{\gamma \in \Gamma} \left(P_{\theta_0,\gamma}(T_n(\theta_0) > c_{1-\alpha/2}) + P_{\theta_0,\gamma}(T_n(\theta_0) < c_{\alpha/2}) \right). \tag{7.2}$$

The asymptotic size of the test is $AsySz(\theta_0) = \limsup_{n\to\infty} ExSz_n(\theta_0)$. The minimum rejection probability, $MinRP_n(\theta_0)$, of the test is the same as $ExSz_n(\theta_0)$ but with "sup" replaced by "inf" and $AsyMinRP(\theta_0) = \liminf_{n\to\infty} MinRP_n(\theta_0)$.

For equal-tailed subsample t tests, we replace Assumptions F1 and F2 by the following assumptions, which are not very restrictive.

Assumption J1. For all $\varepsilon > 0$, $J_{h^0}(c_{h^0}(\tau) + \varepsilon) > \tau$ for $\tau = \alpha/2$ and $\tau = 1 - \alpha/2$, where $c_{h^0}(\tau)$ is the τ quantile of J_{h^0} and h^0 is as in Assumption B1.

Assumption J2. 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.

Define

$$\begin{aligned} Max^{r-}_{ET,Fix}(\alpha) &= \sup_{h \in H}[1 - J_h(c_{Fix}(1-\alpha/2)) + J_h(c_{Fix}(\alpha/2)-)], \\ Max^{\ell-}_{ET,Fix}(\alpha) &= \sup_{h \in H}[1 - J_h(c_{Fix}(1-\alpha/2)-) + J_h(c_{Fix}(\alpha/2))], \\ Max^{r-}_{ET,Sub}(\alpha) &= \sup_{(g,h) \in GH}[1 - J_h(c_g(1-\alpha/2)) + J_h(c_g(\alpha/2)-)], \text{ and} \\ Max^{\ell-}_{ET,Sub}(\alpha) &= \sup_{(g,h) \in GH}[1 - J_h(c_g(1-\alpha/2)-) + J_h(c_g(\alpha/2))]. \end{aligned} \tag{7.3}$$

Here "$r-$" denotes that the limit from the left "$-$" appears in the *right* summand in the expression for $Max^{r-}_{ET,Fix}(\alpha)$. Analogously, "$\ell-$" denotes that it appears in the *left* summand in the expression for $Max^{\ell-}_{ET,Fix}(\alpha)$. Define $Min^{r-}_{ET,Fix}(\alpha), ..., Min^{\ell-}_{ET,Sub}(\alpha)$ analogously with "inf" in place of "sup."

In the "continuous limit" case, $Max^{r-}_{ET,Sub}(\alpha)$ simplifies to $\sup_{h \in H}[1 - J_h(c_h(1 - \alpha/2)) + J_h(c_h(\alpha/2)-)]$ and likewise for $Max^{\ell-}_{ET,Sub}(\alpha)$.

The proofs of Theorems 1 and 2 can be adjusted straightforwardly to yield the following results for equal-tailed FCV and subsample t tests.

Corollary 2 *Let* $\alpha \in (0, 1/2)$ *be given. Let* $T_n(\theta_0)$ *be defined as in Assumption* t1(i).

(a) *Suppose Assumption* B1(i) *holds. Then, an equal-tailed FCV* t *test satisfies*

$$P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{Fix}(1-\alpha/2) \text{ or } T_n(\theta_0) < c_{Fix}(\alpha/2)) \\ \to [1 - J_h(c_{Fix}(1-\alpha/2)) + J_h(c_{Fix}(\alpha/2)-), \\ 1 - J_h(c_{Fix}(1-\alpha/2)-) + J_h(c_{Fix}(\alpha/2))].$$

(b) *Suppose Assumptions* A1, B1, C-E, G1, *and* J1 *hold. Then, an equal-tailed subsample* t *test satisfies*

$$P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{n,b}(1-\alpha/2) \text{ or } T_n(\theta_0) < c_{n,b}(\alpha/2)) \\ \to [1 - J_h(c_{h^0}(1-\alpha/2)) + J_h(c_{h^0}(\alpha/2)-), 1 - J_h(c_{h^0}(1-\alpha/2)-) + J_h(c_{h^0}(\alpha/2))].$$

(c) *Suppose Assumptions* A2 *and* B2 *hold. Then, an equal-tailed FCV* t *test satisfies*

$$AsySz(\theta_0) \in [Max^{r-}_{ET,Fix}(\alpha), Max^{\ell-}_{ET,Fix}(\alpha)] \text{ and} \\ AsyMinRP(\theta_0) \in [Min^{r-}_{ET,Fix}(\alpha), Min^{\ell-}_{ET,Fix}(\alpha)].$$

(d) *Suppose Assumptions* A2, B2, C-E, G2, *and* J2 *hold. Then, an equal-tailed subsample* t *test satisfies the result of part* (c) *with* Sub *in place of* Fix.

Comments. 1. If $J_h(x)$ is continuous at the appropriate value(s) of x, then $Max^{r-}_{ET,Fix}(\alpha) = Max^{\ell-}_{ET,Fix}(\alpha)$ etc. and Corollary 2 gives the precise value of $AsySz(\theta_0)$.

2. By Corollary 2(d) and the definition of $Max^{\ell-}_{ET,Sub}(\alpha)$, sufficient conditions for a nominal level α equal-tailed subsample test to have asymptotic level α are the following: (a) $c_g(1-\alpha/2) \geq c_h(1-\alpha/2)$ for all $(g,h) \in GH$, (b) $c_g(\alpha/2) \leq c_h(\alpha/2)$ for all $(g,h) \in GH$, and (c) $\sup_{h \in H}[1-J_h(c_h(1-\alpha/2)-)+J_h(c_h(\alpha/2))] = \sup_{h \in H}[1-J_h(c_h(1-\alpha/2))+J_h(c_h(\alpha/2)-)]$. Conditions (a) and (b) automatically hold in "continuous limit" cases. They also hold in some "discontinuous limit" cases, but often fail in such cases. Condition (c) holds in most examples. (Note that conditions (a)-(c) are not *necessary* for a subsample test to have asymptotic level α.)

3. Theorems 1 and 2 give results concerning the null rejection rates for each tail separately of an equal-tailed t test. If one is interested in an equal-tailed t test, rather than a symmetric t test, such rates are of interest.

Example 1 (cont.). The critical values $(c_{\alpha/2}, c_{1-\alpha/2})$ for the equal-tailed FCV and subsample tests are $(z_{\alpha/2}, z_{1-\alpha/2})$ and $(c_{n,b}(\alpha/2), c_{n,b}(1-\alpha/2))$, respectively. For $\alpha < 1/2$, Assumption J2 holds by the same sort of argument as used above to verify

Assumption F2. In addition, $Max^{r-}_{ET,Type}(\alpha) = Max^{\ell-}_{ET,Type}(\alpha)$ for $Type = Fix$ and Sub. See the end of the first section of Appendix A for proofs of these claims.

We calculate $Max^{r-}_{ET,Type}(\alpha)$ in (7.3) for $Type = Fix$ and Sub via simulation.

Table I reports the maximum asymptotic null rejection probabilities for the equal-tailed tests given h_2 for a range of h_2 values in $[-1, 1]$ and $\alpha = .05$. (The maximum is over $h \in H$ or $(g, h) \in GH$ with h_2 fixed.) We find a very large size distortion for the equal-tailed subsample test, i.e., $AsySz(\theta_0)$ is about $1/2 + \alpha/2 = .525$, and no size distortion for the two-sided FCV test, i.e., $AsySz(\theta_0)$ is about $\alpha = .05$. But, the two-sided FCV test is not asymptotically similar.

8 Confidence Intervals

In this section, we consider CIs for a parameter $\theta \in R^d$ when nuisance parameters $\eta \in R^s$ and $\gamma_3 \in \mathcal{T}_3$ may appear. To avoid considerable repetition, we recycle the definitions, assumptions, and results given in earlier sections for tests, but with θ and η defined to be part of the vector γ. In previous sections, θ and γ are separate parameters. Here, θ is a sub-vector of γ. The reason for making this change is that the confidence level of a CI for θ by definition depends on uniformity over both the nuisance parameters η and γ_3 *and* the parameter of interest θ. In contrast, the level of a test concerning θ only depends on uniformity over the nuisance parameters and not over θ (because θ is fixed under the null hypothesis). By making θ a sub-vector of γ, the results from previous sections, which are uniform over $\gamma \in \Gamma$, give the uniformity results that we need for CIs for θ. Of course, with this change, the index parameter h, the asymptotic distributions $\{J_h : h \in H\}$, and the assumptions are different in any given model in this CI section from the earlier test sections.

Specifically, we partition θ into $(\theta_1', \theta_2')'$, where $\theta_j \in R^{d_j}$ for $j = 1, 2$, and we partition η into $(\eta_1', \eta_2')'$, where $\eta_j \in R^{s_j}$ for $j = 1, 2$. Then, we consider the same set-up as in Section 5 where $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, but with $\gamma_1 = (\theta_1', \eta_1')'$ and $\gamma_2 = (\theta_2', \eta_2')'$, where $p = d_1 + s_1$ and $q = p_2 + s_2$. Thus, θ and η are partitioned such that θ_1 and η_1 determine whether γ is close to the point of discontinuity of the asymptotic distribution of the test statistic $T_n(\theta)$, whereas θ_2 and η_2 do not, but they still may affect the limit distribution of $T_n(\theta)$. In most examples, either no parameter θ_1 or θ_2 appears (i.e., $d_1 = 0$ or $d_2 = 0$) and either no parameter η_1 or η_2 appears (i.e., $s_1 = 0$ or $s_2 = 0$).

8.1 Basic Results for Confidence Intervals

We consider the same test statistic $T_n(\theta_0)$ for testing the null hypothesis $H_0 : \theta = \theta_0$ as above. Fixed and subsample critical values are defined as above. We obtain CIs for θ by inverting tests based on $T_n(\theta_0)$. When a fixed critical value is employed, this yields an FCV CI. When a subsample critical value is employed, it yields a subsample CI. Let Θ $(\subset R^d)$ denote the parameter space for θ and let Γ denote the parameter space for γ. The CI for θ contains all points $\theta_0 \in \Theta$ for which the test of $H_0 : \theta = \theta_0$

fails to reject the null hypothesis:

$$CI_n = \{\theta_0 \in \Theta : T_n(\theta_0) \leq c_{1-\alpha}\}, \tag{8.1}$$

where $c_{1-\alpha}$ is a critical value equal to $c_{Fix}(1-\alpha)$ or $c_{n,b}(1-\alpha)$.

For example, suppose $T_n(\theta_0)$ is a (i) upper one-sided, (ii) lower one-sided, or (iii) symmetric two-sided t test of nominal level α (i.e., Assumption t1(i), (ii), or (iii) holds). Then, the corresponding CI of nominal level α is defined by

$$\begin{aligned} CI_n &= [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}, \infty), \\ CI_n &= (-\infty, \widehat{\theta}_n + \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}], \text{ or} \\ CI_n &= [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}, \widehat{\theta}_n + \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}], \end{aligned} \tag{8.2}$$

respectively.

The coverage probability of the CI defined in (8.1) when γ is the true parameter vector is

$$P_\gamma(\theta \in CI_n) = P_\gamma(T_n(\theta) \leq c_{1-\alpha}) := 1 - RP_n(\gamma), \tag{8.3}$$

where probabilities are indexed by $\gamma = ((\theta_1', \eta_1')', (\theta_2', \eta_2')', \gamma_3)$ here, whereas they are indexed by (θ, γ) in earlier sections. The exact and asymptotic confidence sizes of CI_n are

$$ExCS_n = \inf_{\gamma \in \Gamma}(1 - RP_n(\gamma)) \text{ and } AsyCS = \liminf_{n\to\infty} ExCS_n, \tag{8.4}$$

respectively. Note that the confidence size depends on uniformity over both θ and η because $\gamma = (\gamma_1, \gamma_2, \gamma_3) = ((\theta_1', \eta_1')', (\theta_1', \eta_1')', \gamma_3)$.

We employ the same assumptions as in Section 5 but with the following changes.

Assumption Adjustments for CIs: (i) θ is a sub-vector of γ, rather than a separate parameter from γ. In particular, $\gamma = (\gamma_1, \gamma_2, \gamma_3) = ((\theta_1', \eta_1')', (\theta_2', \eta_2')', \gamma_3)$ for $\theta = (\theta_1', \theta_2')'$ and $\eta = (\eta_1', \eta_2')'$.

(ii) Instead of the true probabilities under a sequence $\{\gamma_{n,h} : n \geq 1\}$ being $\{P_{\theta_0, \gamma_{n,h}}(\cdot) : n \geq 1\}$, they are $\{P_{\gamma_{n,h}}(\cdot) : n \geq 1\}$.

(iii) The test statistic $T_n(\theta_0)$ is replaced in the assumptions under a true sequence $\{\gamma_{n,h} : n \geq 1\}$ by $T_n(\theta_{n,h})$, where $\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3})' = ((\theta_{n,h,1}', \eta_{n,h,1}')', (\theta_{n,h,2}', \eta_{n,h,2}')', \gamma_{n,h,3})$ and $\theta_{n,h} = (\theta_{n,h,1}', \theta_{n,h,2}')'$.

(iv) In Assumption D, θ_0 in $T_{n,b_n,j}(\theta_0)$ and $T_{b_n}(\theta_0)$ is replaced by θ, where $\theta = (\theta_1', \theta_2')'$ and $\gamma = (\gamma_1, \gamma_2, \gamma_3) = ((\theta_1', \eta_1')', (\theta_2', \eta_2')', \gamma_3)$.

(v) θ_0 is replaced in the definition of $U_{n,b}(x)$ in (3.9) by θ_n when the true parameter is $\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) = ((\theta_{n,1}', \eta_{n,1}')', (\theta_{n,2}', \eta_{n,2}')', \gamma_{n,3})$ and $\theta_n = ((\theta_{n,1}', \theta_{n,2}')'$.

With these changes in the assumptions and corresponding changes in the proofs, the proofs of Theorems 1 and 2 go through.[11] This yields the following results for FCV and subsample CIs.

[11] In the proofs of Corollary 3(c) and (d), $AsySz(\theta_0)$ is replaced by $1 - AsyCS$, $RP_n(\theta_0, \gamma)$ is replaced by $RP_n(\gamma)$, and one makes use of the fact that $\inf_{h\in H} J_h(c_{Fix}(1-\alpha)-) = 1 - Max_{Fix}^-(\alpha)$, $\inf_{(g,h)\in GH} J_h(c_g(1-\alpha)-) = 1 - Max_{Sub}^-(\alpha)$, etc.

Corollary 3 *Let the assumptions be adjusted as stated above.*

(a) *Suppose Assumption* B1(i) *holds. Then,* $P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_{Fix}(1-\alpha)) \to [J_h(c_{Fix}(1-\alpha)-), J_h(c_{Fix}(1-\alpha))]$.

(b) *Suppose Assumptions* A1, B1, C-E, F1*, and* G1 *hold. Then,* $P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_{n,b}(1-\alpha)) \to [J_h(c_{h^0}(1-\alpha)-), J_h(c_{h^0}(1-\alpha))]$.

(c) *Suppose Assumptions* A2 *and* B2 *hold. Then, the* FCV CI *satisfies* $AsyCS \in [\inf_{h\in H} J_h(c_{Fix}(1-\alpha)-), \inf_{h\in H} J_h(c_{Fix}(1-\alpha))]$.

(d) *Suppose Assumptions* A2, B2, C-E, F2*, and* G2 *hold. Then, the subsample* CI *satisfies* $AsyCS \in [\inf_{(g,h)\in GH} J_h(c_g(1-\alpha)-), \inf_{(g,h)\in GH} J_h(c_g(1-\alpha))]$.

Comments. 1. The result of part (a) shows that if $J_h(c_{Fix}(1-\alpha)) < 1-\alpha$ for some $h \in R^{p+q}$, then the FCV CI has asymptotic confidence size less than its nominal level $1-\alpha$: $AsyCS < 1-\alpha$. Similarly, if $J_h(c_{h^0}(1-\alpha)) < 1-\alpha$, then part (b) shows that the subsample CI has asymptotic confidence size less than its nominal level $1-\alpha$: $AsyCS < 1-\alpha$. Parts (c) and (d) establish $AsyCS$ precisely.

2. The lower bound on $AsyCS$ in Corollary 3(d) can be improved in some cases, see the fourth section of Appendix A.

8.2 Equal-tailed t Confidence Intervals

An equal-tailed FCV or subsample t CI for θ of nominal level α is defined by

$$CI_n = [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha/2}, \widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{\alpha/2}], \tag{8.5}$$

where $c_\tau = c_{Fix}(\tau)$ for $\tau = \alpha/2, 1-\alpha/2$ for an FCV CI and $c_\tau = c_{n,b}(\tau)$ for a subsample CI. The following result for such CIs follows from Corollary 2(c)-(d) for equal-tailed t tests, but with the assumptions adjusted as above. For brevity, we do not give analogues of Corollary 2(a)-(b) for CIs.

Corollary 4 *Let* $\alpha \in (0, 1/2)$ *be given. Let the assumptions be adjusted as described above Corollary* 3*.*

(a) *Suppose Assumptions* A2 *and* B2 *hold. Then, an equal-tailed FCV* t *CI satisfies*

$$AsyCS \in [1 - Max^{\ell-}_{ET,Fix}(\alpha), 1 - Max^{r-}_{ET,Fix}(\alpha)].$$

(b) *Suppose Assumptions* A2, B2, C-E, G2*, and* J2 *hold. Then, an equal-tailed subsample* t *CI satisfies the result of part* (a) *with* Sub *in place of* Fix.

9 Examples

9.1 CI Based on a Post-Model-Selection/Super-Efficient Estimator

Here we consider a subsample CI that is based on an estimator that can be viewed either as a post-model-selection estimator based on a consistent model selection procedure or as a super-efficient estimator. We show that the subsample CI (of nominal level

$1-\alpha$ for any $\alpha \in (0,1)$) has $AsyCS = 0$ in a very simple regular model. The same result holds for the $b_n < n$ bootstrap provided $b_n^2/n \to 0$. Similar results hold in much more complicated models than that considered below. Subsampling a super-efficient estimator has been suggested in PRW. Using the $b_n < n$ bootstrap for a post-model-selection estimator has been suggested by Shao (1994).

Kabaila (1995) shows that an FCV CI based on a super-efficient estimator has $AsyCS = 0$, see Leeb and Pötscher (2005) for related results. Hence, we do not consider FCV CIs here. Results of Leeb and Pötscher (2006) show that no uniformly consistent estimator of the distribution of a super-efficient estimator exists. The results below are not a special case of their result, because a uniformly consistent estimator of the null distribution of a test statistic is not necessary to obtain a test of level α; it is sufficient to have an estimator of the null distribution that yields an asymptotically uniformly conservative critical value. Section 9.2 below gives an important example where the latter exists, but the former does not. The literature on post-model-selection inference has been growing recently, see Leeb and Pötscher (2005) for references.

The model is

$$X_i = \theta + U_i, \text{ where } U_i \sim \text{i.i.d. } N(0,1) \text{ for } i = 1, ..., n. \tag{9.1}$$

For the model selection problem, model 1 has $\theta = 0$ and model 2 has $\theta \neq 0$. Model selection is carried out using a likelihood ratio test that selects model 1 if $n^{1/2}|\overline{X}_n| \leq \kappa_n$ and model 2 otherwise, where $\kappa_n > 0$ is a critical value. If $\kappa_n \to \infty$ and $\kappa_n/n^{1/2} \to 0$ as $n \to \infty$, the model selection procedure is consistent. (That is, when $\theta_0 = 0$, model 1 is chosen with probability that goes to one as $n \to \infty$, and when $\theta_0 \neq 0$, model 2 is chosen with probability that goes to one as $n \to \infty$, where θ_0 is fixed and does not depend on n.) For the results that follow we only use the condition $\kappa_n \to \infty$. When $\kappa_n = \sqrt{\log(n)}$, this model selection procedure is BIC. The AIC criterion is not covered by the results given below because it corresponds to $\kappa_n = \sqrt{2} \nrightarrow \infty$. The post-model selection estimator of θ_0 equals zero if model 1 is selected and $\overline{X}_n$ if model 2 is selected. This estimator is a super-efficient estimator whenever $\kappa_n \to \infty$ and $\kappa_n/n^{1/2} \to 0$. It corresponds to Hodges super-efficient estimator when $\kappa_n = n^{1/4}$.

The post-model-selection/super-efficient estimator, $\widehat{\theta}_n$, of θ and the test statistic, $T_n(\theta_0)$, are defined by

$$\begin{aligned} \widehat{\theta}_n &= \begin{cases} \overline{X}_n & \text{if } n^{1/2}|\overline{X}_n| > \kappa_n \\ a\overline{X}_n & \text{if } n^{1/2}|\overline{X}_n| \leq \kappa_n, \end{cases} \quad \text{where } \overline{X}_n = n^{-1}\sum_{i=1}^n X_i, \\ T_n(\theta_0) &= |n^{1/2}(\widehat{\theta}_n - \theta_0)|, \end{aligned} \tag{9.2}$$

$\kappa_n > 0$, and $0 \leq a < 1$. A post-model-selection estimator is obtained by taking $a = 0$. Hodges' super-efficient estimator is obtained by taking $\kappa_n = n^{1/4}$. For a super-efficient estimator, the constant a is a tuning parameter that determines the magnitude of shrinkage. The test statistic is a two-sided non-studentized t statistic, so that Assumptions t1(iii) and t2 hold with $\tau_n = n^{1/2}$.

The CI for θ is given by the third equation in (8.2) with $c_{1-\alpha}$ equal to the subsample critical value $c_{n,b}(1-\alpha)$ based on subsample statistics $\{T_{n,b_n,j}(\widehat{\theta}_n) : j = 1, ..., q_n\}$ defined in equation (iii) of (3.8) with $\widehat{\sigma}_{n,b,j} = 1$. Note that Assumption Sub1 holds. (The results given below also hold if Assumption Sub2 holds.) We take $\{b_n : n \geq 1\}$ so that Assumption C holds.

We apply Corollary 3(b) with $\gamma = \gamma_1 = \theta = \theta_1 \in R$, $p = d = 1$, and $\Gamma = \Gamma_1 = \Theta = R$. (No γ_2, γ_3, θ_2, or η parameters appear in this example.) The assumptions of Corollary 3(b) are verified below. We take $r = 1/2$ and $\gamma_{n,h}$ $(= \theta_{n,h}) = hn^{-1/2}$, where $h \in R$, in Assumption B1. When the true value is $\theta_{n,h}$, $\widehat{\theta}_n = a\overline{X}_n$ wp$\rightarrow$ 1, see (10.5) in Appendix A. Hence, wp$\rightarrow$ 1, we have

$$\begin{aligned} T_n(\theta_{n,h}) &= |n^{1/2}(a\overline{X}_n - \theta_{n,h})| \\ &= |an^{1/2}(\overline{X}_n - \theta_{n,h}) + (a-1)h| \\ &\rightarrow_d |aZ + (a-1)h| \sim J_h, \text{ where } Z \sim N(0,1) \text{ and} \\ J_h(x) &= \begin{cases} \Phi(a^{-1}(x + (1-a)h)) - \Phi(a^{-1}(-x + (1-a)h)) & \text{if } a \in (0,1) \\ 1(x \geq |h|) & \text{if } a = 0 \end{cases}, \end{aligned} \tag{9.3}$$

using the CLT. Given that $p = d = 1$, we have $h^0 = 0$ and $J_{h^0} = J_0$. For $a = 0$, $J_0(x) = 1(x \geq 0)$ and $c_{h^0}(1-\alpha) = c_0(1-\alpha) = 0$. For $a \in (0,1)$, we have $J_0(x) = \Phi(a^{-1}x) - \Phi(-a^{-1}x)$ and $c_{h^0}(1-\alpha) = c_0(1-\alpha) = az_{1-\alpha/2}$.

For $a = 0$, Corollary 3(b) implies that the limit of the coverage probability of the subsample CI under $\gamma_{n,h}$ $(= \theta_{n,h}) = hn^{-1/2}$ is

$$J_h(c_{h^0}(1-\alpha)) = J_h(0) = 1(0 \geq |h|) = 0 \text{ for } |h| > 0. \tag{9.4}$$

Hence, for $a = 0$, $AsyCS = 0$ for the subsample CI.

For $a \in (0,1)$, the limit of the coverage probability of the subsample CI under $\gamma_{n,h}$ $(= \theta_{n,h}) = hn^{-1/2}$ is

$$J_h(c_{h^0}(1-\alpha)) = J_h(az_{1-\alpha/2}). \tag{9.5}$$

Using (9.3), for $a \in (0,1)$, we have

$$\lim_{h\rightarrow\infty} J_h(az_{1-\alpha/2}) = 0. \tag{9.6}$$

Hence, for $a \in (0,1)$ and h sufficiently large, the asymptotic coverage probability of the symmetric two-sided subsample CI is arbitrarily close to zero. Since $h \in R$ is arbitrary, this implies that $AsyCS = 0$ for this CI.

We obtain the same result that $AsyCS = 0$ if one-sided CIs or equal-tailed two-sided CIs are considered. Furthermore, the size-correction methods of Andrews and Guggenberger (2005a) do not work in this example because Assumptions LF, LS, and LH of that paper fail. (For example, Assumption LF fails when $a = 0$ because $H = R_\infty$, $c_h(1-\alpha) = |h|$, and $\sup_{h \in R_\infty} |h| = \infty$.)

It remains to verify Assumptions A1, B1, C-E, F1, and G1 for arbitrary choice of the parameter h. This is done in Appendix A.

9.2 Confidence Region Based on Moment Inequalities

Here we consider a confidence region (CR) for a true parameter θ_0 ($\in \Theta \subset R^d$) that is defined by moment inequalities and equalities. The true value need not be identified. The CR is obtained by inverting tests that are based on a generalized method of moments-type (GMM) criterion function. This method is introduced by Chernozhukov, Hong, and Tamer (2002) (CHT), who use subsampling to obtain a critical value.[12] Romano and Shaikh (2005a) also considers this method and shows that the limit of finite-sample size is the nominal level in one-sample and two-sample mean problems when a subsample critical value is employed.[13] In this section, we show that this result holds quite generally for subsample CRs of this type—no specific assumptions concerning the form of the moment functions are necessary even though the asymptotic distribution of the test statistic is discontinuous (in the sense discussed above). Note that the results given here are for CRs for the true parameter, rather than for the "identified set."

We also consider CRs based on fixed and "plug-in" critical values—defined below. These critical values are bounded above as $n \to \infty$, whereas subsample critical values diverge to infinity at a rate that depends on the subsample size b_n. The plug-in critical values (PCV) lead to more powerful tests and smaller CRs than the fixed critical values (FCV).

The test statistic considered below is similar to those considered by (i) Moon and Schorfheide (2004), who consider an empirical likelihood version of the GMM criterion function and assume identification of θ_0, (ii) Soares (2005), who allows for the plug-in of preliminary estimators in a GMM and/or empirical likelihood criterion function, and (iii) Rosen (2005), who considers a minimum distance version of the test statistic. By similar arguments to those given below, one can show that the limit of the finite-sample size of a subsample CR based on any one of these test statistics equals its nominal level. The results for fixed and plug-in critical values also extend to these test statistics. Hence, for these versions of the test statistics as well, we find that the PCV has the best properties. For brevity, we outline the argument in Appendix A.

The model is as follows. The true value θ_0 ($\in \Theta \subset R^d$) is assumed to satisfy the following moment conditions:

$$\begin{aligned} E_F m_j(W_i, \theta_0) &\geq 0 \text{ for } j = 1, ..., p \text{ and} \\ E_F m_j(W_i, \theta_0) &= 0 \text{ for } j = p+1, ..., p+s, \end{aligned} \tag{9.7}$$

where $\{m_j(\cdot, \theta) : j = 1, ..., p+s\}$ are scalar-valued moment functions and $\{W_i : i \geq 1\}$ are stationary random vectors with joint distribution F.

[12] CHT focus on CRs for the identified set, rather than the true parameter. By definition, the identfied set, Θ_0, is the set of all $\theta \in \Theta$ that satisfy the moment inequalities and equalities when the true value is θ_0. Also, CHT consider a more general criterion function than that considered below. Their asymptotic results do not establish that the limit of the finite sample size of the CR for the true value is the nominal level, which is one of the results shown in this section.

[13] Strictly speaking this is not true because Shaikh's results do not establish uniformity over the true θ value, although it should not be difficult to extend his results to do so.

The sample moment functions are

$$\overline{m}_{n,j}(\theta) = n^{-1}\sum_{i=1}^{n} m_j(W_i,\theta) \text{ for } j = 1,...,p+s. \tag{9.8}$$

The test that we invert to construct a confidence region is based on an Anderson–Rubin-type GMM statistic for testing $H_0 : \theta = \theta_0$ that gives positive weight to the moment inequalities only when they are violated:

$$\begin{aligned} T_n(\theta_0) &= n\sum_{j=1}^{p}(\overline{m}_{n,j}(\theta_0)/\widehat{\sigma}_{n,j}(\theta_0))_-^2 + n\sum_{j=p+1}^{p+s}(\overline{m}_{n,j}(\theta_0)/\widehat{\sigma}_{n,j}(\theta_0))^2, \text{ where} \\ (x)_- &= \begin{cases} x & \text{if } x < 0 \\ 0 & \text{if } x \geq 0 \end{cases} \end{aligned} \tag{9.9}$$

and $\widehat{\sigma}^2_{n,j}(\theta_0)$ is a consistent estimator of $\sigma^2_{F,j}(\theta_0) = \lim_{n\to\infty} Var_F(n^{1/2}\overline{m}_{n,j}(\theta_0))$ for $j = 1,...,p+s$. For example, with i.i.d. observations, one can define $\widehat{\sigma}^2_{n,j}(\theta_0) = n^{-1}\sum_{i=1}^{n} (m_j(W_i,\theta_0) - \overline{m}_{n,j}(\theta_0))^2$.

The subsample statistics are constructed such that Assumption Sub2 holds. (No consistent estimator of the true parameter exists when the latter is unidentified, so a subsample procedure that satisfies Assumption Sub1 is not suitable.)

We now specify $\gamma = (\gamma_1,\gamma_2,\gamma_3)$ for this example. The moment conditions in (9.7) can be written as

$$\begin{aligned} \sigma^{-1}_{F,j}(\theta_0)E_F m_j(W_i,\theta_0) - \gamma_{1,j,0} &= 0 \text{ for } j = 1,...,p \text{ and} \\ \sigma^{-1}_{F,j}(\theta_0)E_F m_j(W_i,\theta_0) &= 0 \text{ for } j = p+1,...,p+s \end{aligned} \tag{9.10}$$

for some $\gamma_{1,0} = (\gamma_{1,1,0},...,\gamma_{1,p,0})' \in R^p_+$. Let $\Omega_0 = \lim_{n\to\infty} Corr_F(n^{1/2}\overline{m}_n(\theta_0))$, where $\overline{m}_n(\theta_0) = (\overline{m}_{n,1}(\theta_0),...,\overline{m}_{n,p+s}(\theta_0))'$ and $Corr_F(n^{1/2}\overline{m}_n(\theta_0))$ denotes the $(p+s)\times(p+s)$ correlation matrix of $n^{1/2}\overline{m}_n(\theta_0)$. Let γ_1, Ω, and θ denote generic parameter values corresponding to the true parameter values $\gamma_{1,0}$, Ω_0, and θ_0, respectively. We take $\gamma = (\gamma_1,\gamma_2,\gamma_3)$ such that $\gamma_1 \in R^p_+$, $\gamma_2 = (\gamma'_{2,1},\gamma'_{2,2})' = (\theta', vech_*(\Omega)')' \in R^q$, where $vech_*(\Omega)$ denotes $vech(\Omega)$ with the diagonal elements of Ω deleted, $q = d + (p+s)(p+s-1)/2$, and $\gamma_3 = F$.

We take $r = 1/2$ and $h = (h_1,h_2)$, where $h_1 \in R^p_{+,\infty}$, $h_2 = (h'_{2,1},h'_{2,2})'$, $h_{2,1} \in$cl(Θ), $h_{2,2} \in$cl$(\Gamma_{2,2})$, and $\Gamma_{2,2}$ is some set of vectors $\gamma_{2,2}$ such that $\gamma_{2,2} = vech_*(C)$ for some $(p+s)\times(p+s)$ correlation matrix C. Hence, $H = R^p_{+,\infty}\times$cl(Θ) $\times$cl$(\Gamma_{2,2})$. Note that h_1 corresponds to γ_1 and, hence, h_1 measures the extent to which the $j = 1,...,p$ moment inequalities deviate from being equalities. Also, $h_{2,1}$ corresponds to θ and $h_{2,2}$ corresponds to $vech_*(\Omega)$.

The parameter spaces for γ_1 and γ_2 are $\Gamma_1 = R^p_+$ and $\Gamma_2 = \Theta\times\Gamma_{2,2}$, respectively. For given $(\gamma_1,\gamma_2) \in \Gamma_1\times\Gamma_2$, the parameter space for F is

$$\begin{aligned} \mathcal{F}(\gamma_1,\gamma_2) = \{F : \sigma^{-1}_{F,j}(\theta)E_F m_j(W_i,\theta) - \gamma_{1,j} = 0 \text{ for } j = 1,...,p, \\ E_F m_j(W_i,\theta) = 0 \text{ for } j = p+1,...,p+s\}, \end{aligned} \tag{9.11}$$

such that $\{\mathcal{F}(\gamma_1, \gamma_2) : (\gamma_1, \gamma_2) \in \Gamma_1 \times \Gamma_2\}$ satisfies the following "convergence condition." By definition, the convergence condition restricts $\{\mathcal{F}(\gamma_1, \gamma_2) : (\gamma_1, \gamma_2) \in \Gamma_1 \times \Gamma_2\}$ so that under any $\{\gamma_{n,h} = (\gamma_{n,h,1}, (\theta'_{n,h}, vech_*(\Omega_{n,h})')', F_{n,h}) : n \geq 1\}$ for any $h \in H$, we have

$$\begin{aligned} &(A_{n,1}, ..., A_{n,p+s})' \to_d Z_{h_{2,2}} \sim N(0, C_{h_{2,2}}) \text{ as } n \to \infty, \text{ where} \\ &A_{n,j} = n^{1/2}(\overline{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}} \overline{m}_{n,j}(\theta_{n,h}))/\sigma_{F_{n,h},j}(\theta_{n,h}), \text{ and} \\ &\widehat{\sigma}_{nj}(\theta_{n,h})/\sigma_{F_{n,h},j}(\theta_{n,h}) \to_p 1 \text{ as } n \to \infty \end{aligned} \tag{9.12}$$

for $j = 1, ..., p+s$, where $C_{h_{2,2}}$ is the $(p+s) \times (p+s)$ correlation matrix for which $vech_*(C_{h_{2,2}}) = h_{2,2}$. For example, if the observations $\{W_i : i \geq 1\}$ are i.i.d. under a fixed γ and $\widehat{\sigma}^2_{nj}(\theta)$ is defined as above, the convergence condition holds if $E_F|m_j(W_i, \theta)|^{2+\delta_1} \leq M$ and $\sigma_{F,j}(\theta) \geq \delta_2$ for $j = 1, ..., p+s$ for some constants $M < \infty$ and $\delta_1, \delta_2 > 0$ that do not depend on F. (This holds by straightforward calculations using the CLT and LLN for i.n.i.d. random vectors that satisfy a uniform $2+\delta_1$ moment bound.) For dependent observations, one needs to specify a specific variance estimator $\widehat{\sigma}^2_{nj}(\theta)$, such as a HAC estimator, before a primitive convergence condition can be stated. For brevity, we do not do so here. Note that (9.12) and the other conditions employed here allow for weak identification because the test statistic considered is an Anderson-Rubin-type statistic.

Given (9.9) and (9.12), Assumption B2 holds because under $\{\gamma_{n,h} : n \geq 1\}$ we have

$$T_n(\theta_{n,h}) \to_d \sum_{j=1}^{p} (Z_{h_{2,2},j} + h_{1,j})^2_- + \sum_{j=p+1}^{p+s} Z^2_{h_{2,2},j} \sim J_h \tag{9.13}$$

for all $h \in H$, where $Z_{h_{2,2}} = (Z_{h_{2,2},1}, ..., Z_{h_{2,2},p+s})'$ and $h_1 = (h_{1,1}, ..., h_{1,p})'$. (Note that $(Z_{h_{2,2},j'} + h_{1,j'})_- = 0$ for any j' in $\{1, ..., p\}$ for which $h_{1,j'} = \infty$.)

For all $(g, h) \in GH$, we have

$$J_g \geq_{ST} J_h, \tag{9.14}$$

where $\geq_{ST}$ denotes "stochastically greater than or equal to." This holds because $\sum_{j=1}^{p}(Z_{h_{2,2},j} + g_{1,j})^2_- \geq \sum_{j=1}^{p}(Z_{h_{2,2},j} + h_{1,j})^2_-$ a.s. for all $0 \leq g_1 \leq h_1$ (where $g_1 = (g_{1,1}, ..., g_{1,p})'$) due to the $(\cdot)_-$ function. Furthermore, when $s > 0$ (i.e., when some equality constraints appear), J_h is continuous at its $1-\alpha$ quantile for all $\alpha < 1/2$. In consequence, when $s > 0$ by Corollary 3(d) and Comment 2 following Theorem 2, for the subsample CR, we have

$$AsyCS = \inf_{(g,h) \in GH} J_h(c_g(1-\alpha)) = 1 - \alpha \tag{9.15}$$

for $\alpha < 1/2$.

For the case when $s = 0$, $J_h(x)$ is discontinuous at $x = 0$ when $h = (\infty^p, h_2)$, where $\infty^p = (\infty, ..., \infty)' \in R^p_\infty$, and this implies that $\inf_{(g,h) \in GH} J_h(c_g(1-\alpha)-) \neq \inf_{(g,h) \in GH} J_h(c_g(1-\alpha))$. Nevertheless, $AsyCS$ equals $1-\alpha$ in this case too. See the fourth section of Appendix A for details. To conclude, in this example, discontinuity of the limit distribution does not cause size distortion for the subsample CR.

We now verify the remaining assumptions needed for Corollary 3. Assumption A2 holds with Γ defined as above. Assumption C holds by choice of b_n. Assumption D holds by stationarity and the standard definition of subsample statistics in the i.i.d. and dependent cases. Assumption E holds by the general argument given in Section 5 for i.i.d. observations and stationary strong mixing observations provided $\sup_{\gamma \in \Gamma} \alpha_\gamma(m) \to 0$ as $m \to \infty$. Assumption F2 holds for all $\alpha < 1/2$ because if $s > 0$, then J_h is absolutely continuous with strictly increasing df, and if $s = 0$, then J_h has support R_+, the df $J_h(x)$ has a jump at $x = 0$ and no other jumps, is strictly increasing on R_+ if any element of h_1 is finite, and equals one on R_+ if all elements of h_1 are infinite. Assumption G2 holds automatically because the subsample procedure satisfies Assumption Sub2.

Next, we discuss FCV CRs. Corollary 3(c), combined with the continuity results concerning J_h given in the discussion of Assumption F2 above and the result above that $J_g \geq_{ST} J_h$ for $(g, h) \in GH$, imply that for an FCV CR

$$AsyCS = \inf_{h \in H} J_h(c_{Fix}(1-\alpha)) = \inf_{h_2 \in H_2} J_{(0,h_2)}(c_{Fix}(1-\alpha)). \tag{9.16}$$

Hence, if $c_{Fix}(1-\alpha)$ is defined such that

$$\inf_{h_2 \in H_2} J_{(0,h_2)}(c_{Fix}(1-\alpha)) = 1 - \alpha, \tag{9.17}$$

then the FCV CR has asymptotic level $1 - \alpha$, as desired. By (9.13), the distribution $J_{(0,h_2)}$ only depends on

$$h_{2,2} = vech_*(\lim_{n \to \infty} Corr_{F_{n,h}}(n^{1/2}\overline{m}_n(\theta_{n,h}))). \tag{9.18}$$

Hence, determination of the value $c_{Fix}(1-\alpha)$ that satisfies (9.17) only requires maximization over the possible asymptotic correlation matrices of $n^{1/2}\overline{m}_n(\theta_{n,h})$. For example if $p = 1$ and $s = 0$, then $J_{(0,h_2)}$ is the distribution of a random variable that is 0 with probability 1/2 and is chi-squared with one degree of freedom with probability 1/2. Hence, no unknown parameter appears. If $p + s = 2$, then $J_{(0,h_2)}$ depends on the scalar $h_{2,2}$, which is the asymptotic correlation between $n^{1/2}\overline{m}_{n,1}(\theta_{n,h})$ and $n^{1/2}\overline{m}_{n,2}(\theta_{n,h})$. For general $p+s$, one can determine $c_{Fix}(1-\alpha)$ such that the infimum in (9.16) equals $1 - \alpha$ via simulation.

Given (9.16), one can design a data-dependent "plug-in" critical value (PCV) that yields a more powerful test than the FCV test and, hence, a smaller CR, because it is closer to being asymptotically similar. Let $c_{Plug}(h_{2,2}, 1-\alpha)$ denote the $1-\alpha$ quantile of $J_{(0,h_2)}(x)$ (which only depends on $h_{2,2}$). Let $\widehat{h}_{2,2,n}$ be a consistent estimator of $h_{2,2}$. The PCV is

$$c_{Plug}(\widehat{h}_{2,2,n}, 1-\alpha), \tag{9.19}$$

where $\widehat{h}_{2,2,n} - h_{2,2} \to_p 0$ and, hence, $c_{Plug}(\widehat{h}_{2,2,n}, 1-\alpha) - c_{Plug}(h_{2,2}, 1-\alpha) \to_p 0$ as $n \to \infty$ under $\{\gamma_{n,h} : n \geq 1\}$. For example, in the case of i.i.d. observations, one can take

$$
\begin{aligned}
\widehat{h}_{2,2,n} &= vech_*\left(\widehat{D}_n^{-1/2}(\theta_0)\widehat{V}_n(\theta_0)\widehat{D}_n^{-1/2}(\theta_0)\right), \text{ where} \\
\widehat{V}_n(\theta_0) &= n^{-1}\sum_{i=1}^{n}(m(W_i,\theta_0)-\overline{m}_n(\theta_0))(m(W_i,\theta_0)-\overline{m}_n(\theta_0))', \\
m(W_i,\theta_0) &= (m_1(W_i,\theta_0),...,m_{p+s}(W_i,\theta_0))', \text{ and} \\
\widehat{D}_n(\theta_0) &= Diag\{\widehat{\sigma}^2_{n,1}(\theta_0),...,\widehat{\sigma}^2_{n,p+s}(\theta_0)\}. \qquad (9.20)
\end{aligned}
$$

The PCV, $c_{Plug}(\widehat{h}_{2,2,n}, 1-\alpha)$, can be computed by simulation.

The use of the PCV yields a CR for the true value θ_0 whose finite-sample size has limit equal to $1-\alpha$ (using (9.16) with Fix replaced by $Plug$). The PCV CR is not asymptotically similar because the limit of its coverage probability exceeds $1-\alpha$ when $h_1 \neq 0$. However, it is closer to being asymptotically similar than the FCV CR is, because $c_{Plug}(h_{2,2}, 1-\alpha)) \leq \sup_{h_2\in H_2} c_{Plug}(h_{2,2}, 1-\alpha) = c_{Fix}(1-\alpha)$ for all $h_{2,2}$ with strict inequality for some $h_{2,2}$.

See Appendix A for some extensions of this example.

9.3 CI for the Endpoint of a Distribution

In this example, we consider an equal-tailed subsample CI for the lower endpoint, θ, of the support of a distribution. The observations $\{X_i : i = 1,...,n\}$ are i.i.d. with distribution F, where F has a density f with respect to Lebesgue measure and support with finite lower endpoint $\theta = F^{-1}(0)$. For example, F could be $U[\theta, 1]$. This example is non-regular in that rate-n consistent estimators of θ are available. Bickel and Freedman (1981) show that the bootstrap is inconsistent in this example (or a slightly different version of it). Several papers have considered the $b < n$ bootstrap in this example and have shown that it is consistent (for any fixed parameter value), see Swanepoel (1986), Shao (1994), and Bickel, Götze, and van Zwet (1997). We show that an equal-tailed subsample CI with nominal level $1-\alpha$ has $AsyCS = 1-\alpha$. Hence, subsampling "works" in this case. This occurs because this is a "continuous limit" example. That is, the asymptotic distribution of the test statistic considered under any fixed parameter is continuous in the parameter. (In addition, one needs continuity of the asymptotic distribution function at the appropriate quantile(s).) This example is of interest in econometrics because it is a special case of production function frontier and some auction models, see Hirano and Porter (2003) and Chernozhukov and Hong (2004).

For brevity, the details of this example are given in Appendix B.

10 Appendix A

10.1 Studentized t Statistics

In this section we provide sufficient conditions for Assumption G2 for the case when T_n is a studentized t statistic and the subsample statistics satisfy Assumption Sub1. This result generalizes Lemma 1 because Assumption t2 is not imposed. The results apply to models with i.i.d., stationary and weakly dependent, or nonstationary observations.

Just as $T_{n,b_n,j}(\theta_0)$ is defined, let $(\widehat{\theta}_{n,b_n,j}, \widehat{\sigma}_{n,b_n,j})$ be the subsample statistics that are defined exactly as $(\widehat{\theta}_n, \widehat{\sigma}_n)$ are defined, but based on the jth subsample of size b_n. In analogy to $U_{n,b_n}(x)$ defined in (3.9), we define

$$U^{\sigma}_{n,b_n}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(d_{b_n} \widehat{\sigma}_{n,b_n,j} \leq x) \qquad (10.1)$$

for a sequence of normalization constants $\{d_n : n \geq 1\}$. Although $U^{\sigma}_{n,b_n}(x)$ depends on $\{d_n : n \geq 1\}$, we suppress the dependence for notational simplicity.

We now state modified versions of Assumptions B2, D, E, and H that are used with studentized statistics when Assumption Sub1 holds.

Assumption BB2. (i) For some $r > 0$, all $h \in H$, all sequences $\{\gamma_{n,h} : n \geq 1\}$, some normalization sequences of positive constants $\{a_n : n \geq 1\}$ and $\{d_n : n \geq 1\}$, 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 $\{\gamma_{n,h} : n \geq 1\}$, (ii) $P_{\theta_0,\gamma_{n,h}}(\widehat{\sigma}_{n,b_n,j} > 0$ for all $j = 1, ..., q_n) \to 1$ under all sequences $\{\gamma_{n,h} : n \geq 1\}$ and all $h \in H$, and (iii) $W_h(0) = 0$ for all $h \in H$.

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

Assumption EE. For all $h \in H$ and all sequences $\{\gamma_{n,h} : n \geq 1\}$ with corresponding normalization $\{d_n : n \geq 1\}$ as in Assumption BB2, $U^{\sigma}_{n,b}(x) - E_{\theta_0,\gamma_{n,h}} U^{\sigma}_{n,b}(x) \to_p 0$ under $\{\gamma_{n,h} : n \geq 1\}$ for all $x \in R$.

Assumption HH. $a_{b_n}/a_n \to 0$.

In most examples, the normalization sequences $\{a_n : n \geq 1\}$ and $\{d_n : n \geq 1\}$ in Assumptions BB2, EE, and HH do not depend on $\{\gamma_{n,h} : n \geq 1\}$. 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 $\{a_n : n \geq 1\}$ and $\{d_n : n \geq 1\}$ need to depend on $\{\gamma_{n,h} : n \geq 1\}$. For example, this occurs when the observations are stationary or nonstationary depending on the value of γ. In particular, it occurs in an autoregressive model with a root that is less than or equal to one, see Andrews and Guggenberger (2005a) for an analysis of this model.

When $\{a_n : n \geq 1\}$ and $\{d_n : n \geq 1\}$ depend on $\{\gamma_{n,h} : n \geq 1\}$, it must be the case that $\tau_n = a_n(\gamma_{n,h})/d_n(\gamma_{n,h})$ does not depend on $\{\gamma_{n,h} : n \geq 1\}$. Also, in this case, Assumption HH becomes: For all sequences $\{\gamma_{n,h} : n \geq 1\}$ for which $b_n^r \gamma_{n,h,1} \to g_1$ for some $g_1 \in R_\infty^p$, $a_{b_n}(\gamma_{n,h})/a_n(\gamma_{n,h}) \to 0$. When d_n depends on $\gamma_{n,h}$, the normalization constant d_{b_n} that appears in $U_{n,b}^{\sigma}(x)$ in Assumption EE is $d_{b_n} = d_{b_n}(\gamma_{n,h})$.

Assumption BB2 implies Assumption B2 with $\tau_n = a_n/d_n$ (by the continuous mapping theorem using Assumption BB2(iii)). Note that there is a certain redundancy of normalization constants in Assumption BB2. If d_n is known (as occurs in some models, but not all models), then without any loss of generality one could absorb d_n into the definition of $\widehat{\sigma}_n$ and take $d_n = 1$ for all n. We do not do this for three reasons. First, if d_n is unknown (because it depends on the unknown true data generating procedure) and is absorbed into $\widehat{\sigma}_n$, then $\widehat{\sigma}_n$ is unknown, which is problematic. Second, if there is a conventional definition of $\widehat{\sigma}_n$, then absorbing d_n into the definition of $\widehat{\sigma}_n$ would preclude its use. Third, it is convenient to keep the assumptions as close as possible to those of PRW.

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 fixed $\gamma \in \Gamma$ or are stationary, strong mixing, and satisfy the condition in (5.9) for each fixed $\gamma \in \Gamma$ provided the subsamples are constructed as described in Section 5 (for the same reason that Assumption E holds in these cases). Assumption HH holds in many examples when Assumption C holds, as is typically the case. However, it does not hold if θ is unidentified when $\gamma = 0$ (because consistent estimation of θ is not possible in this case and $a_n = 1$ in Assumption BB2(i)). For example, this occurs in a model with weak instruments.

The following Lemma generalizes Lemma 1. It does not impose Assumption t2.

Lemma 2 *Assumptions* t1, Sub1, A2, BB2, C, DD, EE, *and* HH *imply Assumption* G2.

Comments. 1. Given Lemma 2, the result of Theorem 2(b) holds for studentized t statistics under Assumptions t1, Sub1, A2, BB2, C, DD, E, EE, F2, and HH. These Assumptions imply Assumptions B2, D, and G2.

2. The proof of Lemma 2 is a variant of the proofs of Theorems 11.3.1(i) and 12.2.2(i) of PRW to allow for nuisance parameters $\{\gamma_{n,h} : n \geq 1\}$ that vary with n and t statistics that may be one- or two-sided.[14]

Example 1 (cont.). Assumption G2 follows from Lemma 2 in this example by noting that Assumptions BB2 and HH hold with $a_n = n^{1/2}$, $d_n = 1$, $\tau_n = n^{1/2}$, $V_h = J_h$,

[14] Lemma 2 does not assume $\tau_{b_n}/\tau_n \to 0$ (only Assumption HH), although PRW's results do. A careful reading of their proof reveals that the assumption $a_{b_n}/a_n \to 0$ is enough to show that $U_{n,b}(x)$ and $L_{n,b}(x)$ have the same probability limits.

and W_h equal to pointmass at one (where, as above, we assume $\sigma_1^2 = 1$ without loss of generality).

We next show that $Max^-_{Type}(\alpha) = Max_{Type}(\alpha)$ for $Type = Fix$ and Sub in this example and verify the formulae given in (6.6). 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

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

$$\begin{aligned} &Max^-_{Sub}(\alpha) \\ &= 1 - \min\{\inf_{(g,h)\in GH} J_h(c_g(1-\alpha)), \inf_{((g_1,-1),(h_1,-1))\in GH} J_{(h_1,-1)}(c_{(g_1,-1)}(1-\alpha)-)\} \\ &= 1 - \inf_{h_2 \in [-1,0]} J_\infty(c_{(0,h_2)}(1-\alpha)). \end{aligned} \tag{10.2}$$

In the second and last equalities of (10.2), 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

$$\begin{aligned} &\inf_{h_1 \in [0,\infty], h_2 \in [-1,0)} J_{(h_1,h_2)}(z_{1-\alpha}-) = J_\infty(z_{1-\alpha}) = 1-\alpha, \\ &\inf_{(g,h)\in GH, h_2\in[0,1]} J_h(c_g(1-\alpha)) \\ &= \min\{\inf_{h_1\in[0,\infty), h_2\in[0,1]} J_{(h_1,h_2)}(c_{(0,h_2)}(1-\alpha)), \inf_{h_1\in[0,\infty], h_2\in[0,1]} J_\infty(c_{(h_1,h_2)}(1-\alpha))\} \\ &= \min\{1-\alpha, 1-\alpha\} = 1-\alpha, \text{ and} \\ &\inf_{((g_1,-1),(h_1,-1))\in GH} J_{(h_1,-1)}(c_{(g_1,-1)}(1-\alpha)-) = J_\infty(c_{(0,-1)}(1-\alpha)). \end{aligned} \tag{10.3}$$

By the same argument as above, $Max_{Fix}(\alpha)$ and $Max_{Sub}(\alpha)$ equal the right-hand side expressions in (10.2). 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 (6.6).

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.

We now verify Assumption J2. 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 J2 holds.

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

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

$$= \alpha/2 + \sup_{h_1 \in [0,\infty]} J_{(h_1,1)}(z_{\alpha/2})$$
$$= \alpha/2 + J_\infty(z_{\alpha/2}) = \alpha, \qquad (10.4)$$

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 $Max_{ET,Fix}^{\ell-}(\alpha) = Max_{ET,Fix}^{r-}(\alpha)$. Similar arguments yield $Max_{ET,Sub}^{\ell-}(\alpha) = Max_{ET,Sub}^{r-}(\alpha)$.

10.2 Examples

10.2.1 CI Based on a Post-Model-Selection/Super-Efficient Estimator (cont.)

Here we verify Assumptions A1, B1, C-E, F1, and G1 for this example for arbitrary choice of the parameter h. Assumption A1 holds because $\Gamma = R$, Assumption C holds by assumption, Assumptions D and E hold because the observations are i.i.d. for each fixed $\theta \in R$, Assumption H holds because $\tau_{b_n}/\tau_n = b_n^{1/2}/n^{1/2} \to 0$ by Assumption C, and Assumption G1 holds by Lemma 1(a) using Assumption H. For $a = 0$, Assumption F1 holds because $J_{h^0}(x) = 1(x \geq 0)$ has a jump at $x = c_{h^0}(1-\alpha) = 0$ with $J_{h^0}(c_{h^0}(1-\alpha)) = 1 > 1 - \alpha$. For $a \in (0,1)$, Assumption F1 holds because $J_{h^0}(x) = \Phi(a^{-1}x) - \Phi(-a^{-1}x)$ is strictly increasing at $c_{h^0}(1-\alpha) = az_{1-\alpha/2}$.

Next, we verify Assumption B1. For any true sequence $\{\gamma_n : n \geq 1\}$ for which $n^{1/2}\gamma_n$ $(= n^{1/2}\theta_n) = O(1)$, we have

$$\begin{aligned} P_{\gamma_n}(n^{1/2}|\overline{X}_n| \leq \kappa_n) &= P_{\gamma_n}(|n^{1/2}(\overline{X}_n - \theta_n) + n^{1/2}\theta_n| \leq \kappa_n) \\ &= P_{\gamma_n}(|O_p(1) + O(1)| \leq \kappa_n) \to 1 \text{ and} \\ P_{\gamma_n}(\widehat{\theta}_n = a\overline{X}_n) &\to 1, \end{aligned} \qquad (10.5)$$

where the second equality uses the fact that $n^{1/2}(\overline{X}_n - \theta_n)$ has mean zero and variance one and the second convergence result uses the definition of $\widehat{\theta}_n$ in (9.2). Therefore (9.3) holds. For the particular sequence $\gamma_{n,h}$ $(= \theta_{n,h}) = hn^{-1/2}$ in Assumption B1(i), (9.3) implies that Assumption B1(i) holds with $J_h(x)$ defined as above. For any sequence $\{\gamma_{n,0} : n \geq 1\}$ as in Assumption B1(ii), we have $n^{1/2}\gamma_{n,0} = O(1)$, (10.5) holds, and (9.3) holds. Hence, Assumption B1(ii) holds with $J_{h^0}(x) = J_0(x)$ defined as above.

10.2.2 Confidence Region Based on Moment Inequalities (cont.)

Here we discuss confidence regions based on moment inequalities using (i) an empirical likelihood criterion function, (ii) a GMM criterion function that employs a consistent preliminary estimator of a vector of identified parameters, (iii) a minimum distance criterion function, and (iv) restrictions on the parameter space $\Gamma_1 = R_+^p$.

For an empirical likelihood-based test statistic, the asymptotic distribution of the test statistic is the same as the GMM-based statistic above. Hence, the same argument as above leads to (9.15) and the subsample, FCV, and PCV CRs have the desired asymptotic level. The PCV CR yields the smallest CR based on an empirical likelihood test statistic.

Next, suppose the population moment functions are of the form $E_F m_j(W_i, \theta_0, \tau_0) \geq 0$ for $j = 1, ..., p$ and $E_F m_j(W_i, \theta_0, \tau_0) = 0$ for $j = p+1, ..., p+s$, where τ_0 is a parameter for which a preliminary asymptotically normal estimator $\widehat{\tau}_n(\theta_0)$ exists, as in Soares (2005). The sample moment functions are of the form $\overline{m}_{n,j}(\theta) = \overline{m}_{n,j}(\theta, \widehat{\tau}_n(\theta))$. In this case, the asymptotic variance of $n^{1/2}\overline{m}_{n,j}(\theta)$, as well as the quantities Ω and $h_{2,2}$, take different values than when τ_0 appears in place of $\widehat{\tau}_n(\theta_0)$. But, the form of the asymptotic distribution given in (9.13) is the same. (This relies on suitable smoothness of $E_F m_j(W_i, \theta_0, \tau_0)$ with respect to τ_0.) In consequence, by the same argument as above, we have $J_g \geq_{ST} J_h$ for $(g, h) \in GH$ and (9.15), (9.16), and the above PCV CR results hold. Hence, use of a preliminary estimator in the GMM criterion function does not cause size distortion for the subsample, FCV, or PCV CRs (provided J_h is properly defined and takes into account the estimation of τ_0 when computing $c_{Fix}(1-\alpha)$ or $c_{Plug}(\widehat{h}_{2,2,n}, 1-\alpha)$).

We now discuss CRs based on a minimum distance test statistic, as in Rosen (2005). (Rosen (2005) does not consider subsample critical values, but we do here.) For testing $H_0 : \theta = \theta_0$, the test statistic is

$$T_n(\theta_0) = \inf_{t=(t_1', 0_s')' : t_1 \in R_+^p} n(\overline{m}_n(\theta_0) - t)' \widehat{V}_n^{-1} (\overline{m}_n(\theta_0) - t), \tag{10.6}$$

where $\overline{m}_n(\theta) = (\overline{m}_{n,1}(\theta), ..., \overline{m}_{n,p+s}(\theta))',$ $\widehat{V}_n$ is a consistent estimator of $V = \lim_{n\to\infty} Var_F(n^{1/2}\overline{m}_n(\theta_0))$ when θ_0 is the true parameter, and V is assumed to be nonsingular. In this case, under $\{\gamma_{n,h} : n \geq 1\}$, some calculations yield

$$\begin{aligned}
&T_n(\theta_{n,h}) \to_d L(h) \sim J_h, \text{ where} \\
&L(h) = \inf_{t_1 \in R_+^p} \left(Z_{h_{2,2}} - \begin{pmatrix} t_1 - h_1 \\ 0_s \end{pmatrix} \right)' C_{h_{2,2}}^{-1} \left(Z_{h_{2,2}} - \begin{pmatrix} t_1 - h_1 \\ 0_s \end{pmatrix} \right) \text{ and} \\
&Z_{h_{2,2}} \sim N(0, C_{h_{2,2}}),
\end{aligned} \tag{10.7}$$

where $C_{h_{2,2}}$ is the correlation matrix defined in (9.12) and $C_{h_{2,2}}$ is assumed to be nonsingular for all $h_{2,2}$ in the parameter space. If $0 \leq g_1 \leq h_1$, then algebra and $R_+^p - g_1 \subset R_+^p - h_1$ give

$$\begin{aligned}
L(h) &= \inf_{t_1^* \in R_+^p - h_1} \left(Z_{h_{2,2}} - \begin{pmatrix} t_1^* \\ 0_s \end{pmatrix} \right)' C_{h_{2,2}}^{-1} \left(Z_{h_{2,2}} - \begin{pmatrix} t_1^* \\ 0_s \end{pmatrix} \right) \\
&\leq \inf_{t_1^* \in R_+^p - g_1} \left(Z_{h_{2,2}} - \begin{pmatrix} t_1^* \\ 0_s \end{pmatrix} \right)' C_{h_{2,2}}^{-1} \left(Z_{h_{2,2}} - \begin{pmatrix} t_1^* \\ 0_s \end{pmatrix} \right) \text{ a.s.}
\end{aligned} \tag{10.8}$$

Also, for $(g,h) \in GH$, we have $0 \le g_1 \le h_1$ and $g_{2,2} = h_{2,2}$. These results imply that $J_g \ge_{ST} J_h$ for $(g,h) \in GH$ and (9.15) holds. Hence, discontinuity of the limit distribution also does not cause size distortion for the CR based on the subsample minimum distance statistic. Analogous results to those above for the FCV and PCV CRs also hold.

All of the discussion above takes the parameter space for $\gamma_1 = (\gamma_{1,1}, ..., \gamma_{1,p})'$ to be $\Gamma_1 = R_+^p$. This is appropriate when a specific value of one component, $\gamma_{1,j}$ say for a $j \in \{1, ..., p\}$, does not put any additional restriction on the other components $\gamma_{1,k}$ for $k \ne j$, but is not appropriate otherwise. For example, consider a location model with interval outcomes. For simplicity, suppose the interval endpoints are integer values. The model is $y_i^* = \theta_0 + u_i$ for a zero mean disturbance u_i and $y_i = [y_i^*]$ for $i = 1, ..., n$, where $[y_i^*]$ denotes the integer part of y_i^*, y_i^* is not observed, and y_i is observed. The interval outcome $[y_i, y_i + 1]$ necessarily includes the unobserved outcome variable y_i^*. Two moment inequalities that place bounds on θ_0 are (i) $-E_{\theta_0} y_i + \theta_0 \ge 0$ and (ii) $E_{\theta_0} y_i + 1 - \theta_0 \ge 0$. In this example, $\gamma_{1,1} = -E_{\theta_0} y_i + \theta_0$ and $\gamma_{1,2} = E_{\theta_0} y_i + 1 - \theta_0$ satisfy the additional restriction $\gamma_{1,1} + \gamma_{1,2} = 1$.

Analysis of the interval outcome model can be done using the general results of this paper as follows. We have $(-E_{\theta_0} y_i + \theta_0) \in [0,1]$. We treat the two cases (a) $(-E_{\theta_0} y_i + \theta_0) \in [0, 1/2]$ and (b) $(-E_{\theta_0} y_i + \theta_0) \in (1/2, 1]$ separately because the asymptotic distribution of $T_n(\theta_{n,h})$ is discontinuous both at $-E_{\theta_0} y_i + \theta_0 = 0$ and at $-E_{\theta_0} y_i + \theta_0 = 1$. For case (a), we define γ_1 via $(-E_{\theta_0} y_i + \theta_0) + \gamma_1 = 0$ for $\gamma_1 \in [0, 1/2]$ and, in consequence, $(E_{\theta_0} y_i + 1 - \theta_0) - (1 - \gamma_1) = 0$. Using these equalities in place of the equalities in (9.10), we can analyze this model in the same way as above with $p = 2$ and $s = 0$. We obtain the same result as above that the limit of finite-sample size is the nominal level for subsample CRs when $(-E_{\theta_0} y_i + \theta_0) \in [0, 1/2]$. For case (b), we define γ_1 via $(E_{\theta_0} y_i + 1 - \theta_0) - \gamma_1 = 0$ for $\gamma_1 \in [0, 1/2]$ and, in consequence, $(-E_{\theta_0} y_i + \theta_0) + (1 - \gamma_1) = 0$. Analogously, using these equalities in place of the equalities in (9.10), we can analyze this model in the same way as above. We obtain the same result as above that the limit of finite-sample size is the nominal level for subsample CRs when $(-E_{\theta_0} y_i + \theta_0) \in (1/2, 1]$. Combining the results from cases (a) and (b) gives the same result for the model in which $(-E_{\theta_0} y_i + \theta_0) \in [0, 1]$.

10.3 Proofs

The following Lemmas are used in the proofs of Theorems 1 and 2. (The expressions $\kappa_n \to [\kappa_{1,\infty}, \kappa_{2,\infty}]$ and $G(x-)$ used below are defined in Section 5.2.)

Lemma 3 *Suppose* (i) *for some df's* $L_n(\cdot)$ *and* $G_L(\cdot)$ *on* R, $L_n(x) \to_p G_L(x)$ *for all* $x \in C(G_L)$, (ii) $T_n \to_d G_T$, *where* T_n *is a scalar random variable and* G_T *is some distribution on* R, *and* (iii) *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* $c_n := \inf\{x \in R : L_n(x) \ge 1 - \alpha\}$, (a) $c_n \to_p c_\infty$ *and* (b) $P(T_n \le c_n) \to [G_T(c_\infty-), G_T(c_\infty)]$.

Comments. 1. Condition (iii) holds if $G_L(x)$ is strictly increasing at $x = c_\infty$ or if $G_L(x)$ has a jump at $x = c_\infty$ with $G_L(c_\infty) > 1-\alpha$ and $G_L(c_\infty -) < 1-\alpha$.

2. If $G_T(x)$ is continuous at c_∞, then the result of part (b) is $P(T_n \leq c_n) \to G_T(c_\infty)$.

Lemma 4 *Suppose Assumptions* A2, B2, C-E, F2, *and* G2 *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* (i) $w_n^r \gamma_{w_n,1} \to h_1$ *for some* $h_1 \in R_\infty^p$, (ii) $b_{w_n}^r \gamma_{w_n,1} \to g_1$ *for some* $g_1 \in R_\infty^p$, *and* (iii) $\gamma_{w_n,2} \to h_2$ *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*

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

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

(c) $U_{w_n,b_{w_n}}(x) \to_p J_g(x)$ *for all* $x \in C(J_g)$ *under* $\{\gamma_{w_n} : n \geq 1\}$,

(d) $L_{w_n,b_{w_n}}(x) \to_p J_g(x)$ *for all* $x \in C(J_g)$ *under* $\{\gamma_{w_n} : n \geq 1\}$,

(e) $c_{w_n,b_{w_n}}(1-\alpha) \to_p c_g(1-\alpha)$ *under* $\{\gamma_{w_n} : n \geq 1\}$,

(f) $P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq c_{w_n,b_{w_n}}(1-\alpha)) \to [J_h(c_g(1-\alpha)-), J_h(c_g(1-\alpha))]$, *and*

(g) *if* $||h_1|| < \infty$ *and* $w_n = n$ *for all* $n \geq 1$, *then parts* (b)-(f) *hold with Assumptions* A2, B2, F2, *and* G2 *replaced by Assumptions* A1, B1, F1, *and* G1 *and* (g, h) *in parts* (b)-(f) *equal* (h^0, h) *of Assumption* B1.

Comment. If J_h is continuous at $c_g(1-\alpha)$, $P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq c_{w_n,b_{w_n}}(1-\alpha)) \to J_h(c_g(1-\alpha))$.

Lemma 5 *Suppose Assumptions* A2, B2, C-E, F2, *and* G2 *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* (i)-(iii) *of Lemma* 4 *and for this sequence parts* (b)-(f) *of Lemma* 4 *hold with* w_n *replaced by* n.

Proof of Lemma 3. For $\varepsilon > 0$ such that $c_\infty \pm \varepsilon \in C(G_L) \cap C(G_T)$, we have

$$\begin{aligned} &L_n(c_\infty - \varepsilon) \to_p G_L(c_\infty - \varepsilon) < 1-\alpha \text{ and} \\ &L_n(c_\infty + \varepsilon) \to_p G_L(c_\infty + \varepsilon) > 1-\alpha \end{aligned} \tag{10.9}$$

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

$$P(A_n(\varepsilon)) \to 1, \text{ where } A_n(\varepsilon) = \{c_\infty - \varepsilon \leq c_n \leq c_\infty + \varepsilon\}. \tag{10.10}$$

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 (a) holds.

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

$$P(T_n \leq c_\infty - \varepsilon, A_n(\varepsilon)) \leq P(T_n \leq c_n, A_n(\varepsilon)) \leq P(T_n \leq c_\infty + \varepsilon). \tag{10.11}$$

Hence,

$$
\begin{aligned}
&\limsup_{n\to\infty} P(T_n \le c_n) = \limsup_{n\to\infty} P(T_n \le c_n, A_n(\varepsilon)) \\
\le &\limsup_{n\to\infty} P(T_n \le c_\infty + \varepsilon) = G_T(c_\infty + \varepsilon), \text{ and} \\
&\liminf_{n\to\infty} P(T_n \le c_n) = \liminf_{n\to\infty} P(T_n \le c_n, A_n(\varepsilon)) \\
\ge &\liminf_{n\to\infty} P(T_n \le c_\infty - \varepsilon, A_n(\varepsilon)) = G_T(c_\infty - \varepsilon) \qquad (10.12)
\end{aligned}
$$

using assumption (ii), $c_\infty \pm \varepsilon \in C(G_T)$, and (10.10). Given a sequence $\{\varepsilon_k : k \ge 1\}$ as above, (10.12) establishes part (b). □

Proof of Lemma 4. First, we prove part (a). We need to show that $g \in H$, $h \in H$, $g_2 = h_2$, and conditions (i)-(iii) in the definition of GH hold. For $m = 1, ..., p$, if $a_m = 0$, then $g_{1,m}, h_{1,m} \in R_{+,\infty}$ by conditions (i) and (ii) of the Lemma. Likewise, if $b_m = 0$, 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 (iii) 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 (i) and (ii) of the Lemma and Assumption C(ii), conditions (i)-(iii) of GH hold. Hence, $(g, h) \in GH$.

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

$$
\begin{aligned}
E_{\gamma_{w_n}} U_{w_n,b_{w_n}}(x) &= q_{w_n}^{-1} \sum_{j=1}^{q_{w_n}} P_{\gamma_{w_n}}(T_{w_n,b_{w_n},j}(\theta_0) \le x) \\
&= P_{\gamma_{w_n}}(T_{w_n,b_{w_n},1}(\theta_0) \le x) = P_{\gamma_{w_n}}(T_{b_{w_n}}(\theta_0) \le x), \qquad (10.13)
\end{aligned}
$$

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 B2 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 (i′) $u_n^r \gamma^*_{u_n,1} \to g_1$ and (ii′) $\gamma^*_{u_n,2} \to g_2 \in R^q$, we have

$$
P_{\gamma^*_{u_n}}(T_{u_n}(\theta_0) \le y) \to J_g(y), \qquad (10.14)
$$

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 (i′) and (ii′) hold by assumptions (ii) and (iii) on $\{\gamma_{w_n} : n \ge 1\}$.

For the proof of part (b), it remains to show (10.14). Because $g \in H$, by definition of H there exists a sequence $\{\gamma_k^+ = (\gamma_{k,1}^+, \gamma_{k,2}^+, \gamma_{k,3}^+) \in \Gamma : k \geq 1\}$ 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 $\{\gamma_k^{**} = (\gamma_{k,1}^{**}, \gamma_{k,2}^{**}, \gamma_{k,3}^{**}) \in \Gamma : k \geq 1\}$ as follows. If $k = u_n$ set γ_k^{**} equal to $\gamma_{u_n}^*$. If $k \neq u_n$, set γ_k^{**} equal to γ_k^+. Clearly, $\gamma_k^{**} \in \Gamma$ for all $k \geq 1$ and $k^r \gamma_{k,1}^{**} \to g_1$ and $\gamma_{k,2}^{**} \to g_2$ as $k \to \infty$. Hence, $\{\gamma_k^{**} : k \geq 1\}$ is of the form $\{\gamma_{n,g} : n \geq 1\}$ and Assumption B2 implies that $P_{\gamma_k^{**}}(T_k(\theta_0) \leq y) \to J_g(y)$ 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 $P_{\gamma_{u_n}^*}(T_{u_n}(\theta_0) \leq y) \to J_g(y)$, as desired.

For part (c) 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 $\{\gamma_{w_n} : n \geq 1\}$. Define a new sequence $\{\gamma_k^* = (\gamma_{k,1}^*, \gamma_{k,2}^*, \gamma_{k,3}^*) \in \Gamma : k \geq 1\}$ as follows. If $k = w_n$, set γ_k^* equal to γ_{w_n}. If $k \neq w_n$, for $m = 1, ..., p$, define

$$\begin{array}{ll}
\gamma_{k,1,m}^* = \max\{k^{-r}h_{1,m}, a_m/2\} & \text{if } g_{1,m} = 0 \;\&\; -\infty < h_{1,m} < 0 \\
\gamma_{k,1,m}^* = \min\{k^{-r}h_{1,m}, b_m/2\} & \text{if } g_{1,m} = 0 \;\&\; 0 < h_{1,m} < \infty \\
\gamma_{k,1,m}^* = \max\{-k^{-2r}, a_m/2\} & \text{if } g_{1,m} = h_{1,m} = 0 \;\&\; a_m < 0 \\
\gamma_{k,1,m}^* = \min\{k^{-2r}, b_m/2\} & \text{if } g_{1,m} = h_{1,m} = 0,\; a_m = 0,\; \&\; b_m > 0 \\
\gamma_{k,1,m}^* = \max\{-(b_k k)^{-r/2}, a_m/2\} & \text{if } g_{1,m} = 0 \;\&\; h_{1,m} = -\infty \\
\gamma_{k,1,m}^* = \min\{(b_k k)^{-r/2}, b_m/2\} & \text{if } g_{1,m} = 0 \;\&\; h_{1,m} = \infty \\
\gamma_{k,1,m}^* = \max\{b_k^{-r} g_{1,m}, a_m/2\} & \text{if } -\infty < g_{1,m} < 0 \;\&\; h_{1,m} = -\infty \\
\gamma_{k,1,m}^* = \min\{b_k^{-r} g_{1,m}, b_m/2\} & \text{if } 0 < g_{1,m} < \infty \;\&\; h_{1,m} = \infty \\
\gamma_{k,1,m}^* = a_m/2 & \text{if } g_{1,m} = h_{1,m} = -\infty \\
\gamma_{k,1,m}^* = b_m/2 & \text{if } g_{1,m} = h_{1,m} = \infty,
\end{array} \tag{10.15}$$

where $\gamma_{k,1}^* = (\gamma_{k,1,1}^*, ..., \gamma_{k,1,p}^*)'$, 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 A2(ii) and straightforward calculations show that $\{\gamma_k^* : k \geq 1\}$ satisfies (i)-(iii) of Lemma 4 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 (c) then follows from part (b).

To prove part (d), we show that Assumptions A2 and G2 imply that

$$L_{w_n,b_{w_n}}(x) - U_{w_n,b_{w_n}}(x) \to_p 0 \text{ under } \{\gamma_{w_n} : n \geq 1\} \text{ for all } x \in C(J_g). \tag{10.16}$$

This and part (c) of the Lemma establish part (d). To show (10.16), define the same sequence $\{\gamma_k^*\}$ as in part (c) that satisfies (i)-(iii) of Lemma 4 with $\{w_n\}$ replaced by $\{k\}$. Hence, by Lemma 4(c) 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 G2 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 (10.16) holds.

Parts (e) and (f) are established by applying Lemma 3 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 3 using (I) part (d), (II)

$T_{w_n}(\theta_0) \to_d J_h$ under $\{\gamma_{w_n} : n \geq 1\}$ (which is verified below), and (III) Assumption F2. The result of (II) holds because $\{\gamma_k^* : k \geq 1\}$ in the proof of part (c) is of the form $\{\gamma_{n,h} : n \geq 1\}$ for h as defined in the statement of Lemma 4; this and Assumption B2 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).

Part (g) holds because (I) the proof of part (b) goes through with Assumptions A2 and B2 replaced by Assumptions A1 and B1 given that $||h_1|| < \infty$, which implies that $g = h^0$, (II) the proof of part (c) holds without change, (III) part (d) holds immediately by part (c) and Assumption G1 (in place of Assumption G2) because $w_n = n$ for all $n \geq 1$, and (IV) the proof of parts (e) and (f) holds with Assumptions A2, B2, and F2 replaced by Assumptions A1, B1(i), and F1 given that $||h_1|| < \infty$ (which implies that $g_1 = 0$ using conditions (i) and (ii) of the Lemma and Assumption C(ii)) and $w_n = n$. $\square$

Proof of Lemma 5. Define $\gamma_{n,1,m}$ as in (10.15) with n in place of k for $m = 1, ..., p$ and let $\gamma_{n,1} = (\gamma_{n,1,1}, ..., \gamma_{n,1,p})'$. 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 A2. Also, using Assumption C, straightforward calculations show that $\{\gamma_n : n \geq 1\}$ satisfies conditions (i)-(iii) of Lemma 4 with $w_n = n$. Hence, parts (b)-(f) of Lemma 4 hold with $w_n = n$ for $\{\gamma_n : n \geq 1\}$ as defined above. $\square$

Proof of Theorem 1. Part (a) holds by Assumption B1(i) and the definition of convergence in distribution by considering points of continuity of $J_h(\cdot)$ that are greater than $c_{Fix}(1-\alpha)$ and arbitrarily close to $c_{Fix}(1-\alpha)$ as well as continuity points that are less than $c_{Fix}(1-\alpha)$ and arbitrarily close to it. Part (b) follows from Lemma 4(g) because $||h_1|| < \infty$, $w_n = n$ for all $n \geq 1$, g in Lemma 4(g) equals $h^0 = (0, h_2)$, conditions (i) and (iii) of Lemma 4 hold by the definition of the sequence $\{\gamma_{n,h} : n \geq 1\}$, and condition (ii) of Lemma 4 holds because $n^r \gamma_{n,h,1} \to h_1$ with $||h_1|| < \infty$ implies that $b_n^r \gamma_{n,h,1} \to 0$ using Assumption C(ii). $\square$

Proof of Theorem 2. The proof of part (a) is similar to that of part (b), but noticeably simpler because $c_{Fix}(1-\alpha)$ is a constant. Furthermore, the proof of the second result of part (b) is quite similar to that of the first result. Hence, for brevity, we only prove the first result of part (b).

We first show that $AsySz(\theta_0) \geq Max_{Sub}(\alpha)$. Equations (3.10) and (3.11) imply that for any sequence $\{\gamma_n \in \Gamma : n \geq 1\}$,

$$AsySz(\theta_0) \geq \limsup_{n\to\infty} [1 - P_{\theta_0,\gamma_n}(T_n(\theta_0) \leq c_{n,b}(1-\alpha))]. \tag{10.17}$$

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

$$\limsup_{n\to\infty} [1 - P_{\theta_0,\gamma_n}(T_n(\theta_0) \leq c_{n,b}(1-\alpha))] \geq 1 - J_h(c_g(1-\alpha)). \tag{10.18}$$

The latter inequality holds by Lemma 5.

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

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

$$\begin{aligned} b^r_{w_n} \gamma^*_{w_n,1,m} &\to 0 \text{ and} \\ w^r_n \gamma^*_{w_n,1,m} &\to h_{1,m} \text{ for some } h_{1,m} \in R. \end{aligned} \tag{10.19}$$

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

$$\begin{aligned} b^r_{w_n} \gamma^*_{w_n,1,m} &\to g_{1,m} \text{ for some } g_{1,m} \in R \text{ and} \\ w^r_n \gamma^*_{w_n,1,m} &\to h_{1,m}, \text{ where } h_{1,m} = \infty \text{ or } -\infty \text{ with } sgn(h_{1,m}) = sgn(g_{1,m}). \end{aligned} \tag{10.20}$$

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

$$\begin{aligned} b^r_{w_n} \gamma^*_{w_n,1,m} &\to g_{1,m}, \text{ where } g_{1,m} = \infty \text{ or } -\infty, \text{ and} \\ w^r_n \gamma^*_{w_n,1,m} &\to h_{1,m}, \text{ where } h_{1,m} = \infty \text{ or } -\infty \text{ with } sgn(h_{1,m}) = sgn(g_{1,m}). \end{aligned} \tag{10.21}$$

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

$$\gamma^*_{w_n,2} \to h_2 \text{ for some } h_2 \in \text{cl}(\Gamma_2). \tag{10.22}$$

By taking successive subsequences over the p components of $\gamma^*_{v_n,1}$ and $\gamma^*_{v_n,2}$, we find that there exists a subsequence $\{w_n\}$ of $\{v_n\}$ such that for each $m = 1, ..., p$ exactly one of the cases (10.19)-(10.21) applies and (10.22) holds. In consequence, conditions (i)-(iii) of Lemma 4 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,

$$RP_{w_n}(\theta_0, \gamma^*_{w_n}) \to [1 - J_h(c_g(1-\alpha)), 1 - J_h(c_g(1-\alpha)-)] \tag{10.23}$$

by Lemma 4(f). Also, $(g, h) \in GH$ by Lemma 4(a). Since $\lim_{n\to\infty} RP_{v_n}(\theta_0, \gamma^*_{v_n}) = AsySz(\theta_0)$ and $\{w_n\}$ is a subsequence of $\{v_n\}$, we have $\lim_{n\to\infty} RP_{w_n}(\theta_0, \gamma^*_{w_n}) = AsySz(\theta_0)$. This, (10.23) and $(g, h) \in GH$ imply that $AsySz(\theta_0) \leq Max^{-}_{Sub}(\alpha)$, which completes the proof of the first result of part (b). □

Proof of Lemma 2. 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 $b^r_n \gamma_{n,h,1} \to g_1$ 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 $R_n(t) := q_n^{-1} \sum_{j=1}^{q_n} 1(|\tau_{b_n}(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_{n,b_n,j}| \geq t)$. Using

$$U_{n,b}(x-t) - R_n(t) \leq L_{n,b}(x) \leq U_{n,b}(x+t) + R_n(t) \tag{10.24}$$

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 $R_n(t) \rightarrow_p 0$ under $\{\gamma_{n,h}\}$ for any fixed $t > 0$. By $\tau_n = a_n/d_n$, we have

$$|\tau_{b_n}(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_{n,b_n,j}| \geq t \text{ iff } (a_{b_n}/a_n)a_n|\widehat{\theta}_n - \theta_0| \geq d_{b_n}\widehat{\sigma}_{n,b_n,j}t \tag{10.25}$$

provided $\widehat{\sigma}_{n,b_n,j} > 0$, which by Assumption BB2(ii) holds uniformly in $j = 1, ..., q_n$ wp→1. (In the case where a_n and d_n depend on $\gamma_{n,h}$, the expression on the rhs of (10.25) is $(a_{b_n}(\gamma_{n,h})/a_n(\gamma_{n,h}))a_n(\gamma_{n,h})|\widehat{\theta}_n - \theta_0| \geq d_{b_n}(\gamma_{n,h})\widehat{\sigma}_{n,b_n,j}t$.) By Assumption BB2(i) and HH, $(a_{b_n}/a_n)a_n|\widehat{\theta}_n - \theta_0| = o_p(1)$ under $\{\gamma_{n,h}\}$. Therefore, for any $\delta > 0$, $R_n(t) \leq q_n^{-1}\sum_{j=1}^{q_n} 1(\delta \geq d_{b_n}\widehat{\sigma}_{n,b_n,j}t) = U^{\sigma}_{n,b_n}(\delta/t)$ where the inequality holds wp→1. Now, by an argument as in the proof of Lemma 4(b) and (c) (which uses Assumption EE, but does not use Assumption G2) applied to the statistic $d_n\widehat{\sigma}_n$ rather than $T_{w_n}(\theta_0)$, we have $U^{\sigma}_{n,b_n}(x) \rightarrow_p W_g(x)$ for all $x \in C(W_g)$ under $\{\gamma_{n,h}\}$, where $g \in H$ is defined as in Lemma 4 with $\{\gamma_{w_n}\}$ being equal to $\{\gamma_{n,h}\}$. Therefore, $U^{\sigma}_{n,b_n}(\delta/t) \rightarrow_p W_g(\delta/t)$ for $\delta/t \in C(W_g)$ under $\{\gamma_{n,h}\}$. By Assumption BB2(iii), W_g does not have positive mass at zero and, hence, $W_g(\delta/t) \rightarrow 0$ as $\delta \rightarrow 0$. We can therefore establish that $R_n(t) \rightarrow_p 0$ for any $t > 0$ by letting δ go to zero such that $\delta/t \in C(W_g)$. □

10.4 Additional Results

The upper bound, $Max^-_{Sub}(\alpha)$, on $AsySz(\theta_0)$ in Theorem 2(b) can be improved in certain situations. For example, this occurs in some models in which $J_h(x)$ has a discontinuity at $x = c_g(1-\alpha)$ for some $h \in H$ and the test statistic and the subsample statistics have a common lower bound on their support for all $n \geq 1$. The improvement is possible because the test statistic and the subsample critical values cannot be smaller than the lower bound. For example, this improvement is relevant in the moment inequality example when $s = 0$ (i.e., when no equality constraints appear).

Modification to Theorem 2(b): Let GH^* be a set of points $(g,h) \in GH$ such that for all sequences $\{\gamma_{w_n} : n \geq 1\}$ that satisfy (i)-(iii) of Lemma 4, we have

$$\liminf_{n\rightarrow\infty} P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq c_{w_n,b_{w_n}}(1-\alpha)) \geq J_h(c_g(1-\alpha)). \tag{10.26}$$

Then, $Max^-_{Sub}(\alpha)$ in Theorem 2(b) can be replaced by $Max^-_{Sub,2}(\alpha)$, which is defined by

$$\max\left\{\sup_{(g,h)\in GH\setminus GH^*}(1 - J_h(c_g(1-\alpha)-)), \sup_{(g,h)\in GH^*}(1 - J_h(c_g(1-\alpha)))\right\}. \tag{10.27}$$

Clearly, $Max^{-}_{Sub,2}(\alpha) \leq Max^{-}_{Sub}(\alpha)$.

Analogously, if $(h^0, h) \in GH^*$ in Theorem 1, then $J_h(c_{h^0}(1-\alpha)-)$ can be replaced by $J_h(c_{h^0}(1-\alpha))$ in the result of Theorem 1(b). Similarly, for CIs or CRs, the lower bound in Corollary 3(d), $\inf_{(g,h)\in GH} J_h(c_g(1-\alpha)-)$, can be replaced by

$$\min\left\{\inf_{(g,h)\in GH\backslash GH^*} J_h(c_g(1-\alpha)-),\ \inf_{(g,h)\in GH^*} J_h(c_g(1-\alpha))\right\}, \tag{10.28}$$

where GH^* is defined as above, but with $\theta_{n,h}$ in place of θ_0 in condition (iii), where $\theta_{n,h}$ is a subvector of $\gamma_{n,h}$.

Sufficient conditions for (g, h) to be in GH^* are that for all sequences $\{\gamma_{w_n} : n \geq 1\}$ that satisfy (i)-(iii) of Lemma 4, (a) there exists a finite non-stochastic lower bound LB_h such that the subsample statistics are $\geq LB_h$ a.s. under $\{\gamma_{w_n} : n \geq 1\}$, (b) $J_h(LB_h) \geq J_h(c_g(1-\alpha))$, and (c) $\liminf_{n\to\infty} P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq LB_h) = J_h(LB_h)$. (Conditions (a)-(c) imply (10.26) because $\liminf_{n\to\infty} P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq c_{w_n,b_{w_n}}(1-\alpha)) \geq \liminf_{n\to\infty} P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq LB_h) = J_h(LB_h) \geq J_h(c_g(1-\alpha))$.)

We now use the results above to show that the subsample CR for the moment inequality example of Section 9.2 has $AsyCS = 1-\alpha$ when $s = 0$. In this case, we claim that $GH^* = GH$. Suppose $c_g(1-\alpha) = 0$, then $(g,h) \in GH^*$ because condition (a) of the previous paragraph holds with $LB_h = 0$, condition (b) holds because $J_h(LB_h) = J_h(0) = J_h(c_g(1-\alpha))$, and condition (c) holds by the following argument. First, $c_g(1-\alpha) = 0$ implies that $h_1 = \infty^p$. For notational simplicity, suppose $p = 1$ and let (θ, P, E) denote $(\theta_{w_n,h}, P_{\theta_0,\gamma_{w_n}}, E_{\theta_0,\gamma_{w_n}})$. For the case $s = 0$, we have

$$\begin{aligned}
& P(T_{w_n}(\theta) > LB_h) \\
=\ & P(\min\{w_n^{1/2}\overline{m}_{w_n}(\theta)/\widehat{\sigma}_{w_n}(\theta), 0\}^2 > 0) \\
=\ & P(w_n^{1/2}\overline{m}_{w_n}(\theta) < 0) \\
=\ & P(w_n^{1/2}(\overline{m}_{w_n}(\theta) - E\overline{m}_{w_n}(\theta))/\sigma_{F_{w_n,h},j}(\theta) + w_n^{1/2}E\overline{m}_{w_n}(\theta)/\sigma_{F_{w_n,h},j}(\theta) < 0) \\
\to\ & 0 = 1 - J_h(LB_h),
\end{aligned} \tag{10.29}$$

where the convergence holds because the first summand on the fourth line is $O_p(1)$ and the second diverges to ∞ by (9.10), (9.12), and $h_1 = \infty$. Hence, condition (c) holds. The proof when $p > 1$ is analogous.

Next, suppose $c_g(1-\alpha) > 0$, then $(g,h) \in GH^*$ because $J_h(x)$ is continuous for all $x > 0$ for all $h \in H$, which implies that $J_h(c_g(1-\alpha)-) = J_h(c_g(1-\alpha))$, and then Lemma 4(f) yields (10.26). Now, $GH^* = GH$, (10.28), and Corollary 3(d) imply that $AsyCS = \inf_{(g,h)\in GH} J_h(c_g(1-\alpha))$. Because $J_g \geq_{ST} J_h$ for $(g,h) \in GH$ by (9.14), $\inf_{(g,h)\in GH} J_h(c_g(1-\alpha)) = J_h(c_h(1-\alpha)) = 1-\alpha$ as desired.

Proof of Modification of Theorem 2(b). If we add the assumption that $\liminf_{n\to\infty} P(T_n \leq c_n) \geq G_T(c_\infty)$ in Lemma 3, then the Lemma yields the stronger conclusion that $P(T_n \leq c_n) \to G_T(c_\infty)$. This follows directly from equation (10.12) in the proof of Lemma 3. Therefore, for any $(g, h) \in GH^*$ and sequence $\{\gamma_{w_n} : n \geq 1\}$ that satisfies (i)-(iii) of Lemma 4, the proof of Lemma 4(f) yields the stronger conclusion that $P_{\theta_0,\gamma_{w_n}}(T_{w_n}(\theta_0) \leq c_{w_n,b_{w_n}}(1-\alpha)) \to J_h(c_g(1-\alpha))$. Combining this with the proof of Theorem 2(b) establishes the claim. $\square$

11 Appendix B

11.1 Proofs

We now give a proof of Corollary 1. The purpose is to provide a direct proof of Corollary 1 that is simpler and more transparent than the proof of the more general Theorem 2. The proof uses the following two lemmas. Part (b) of Lemma 6 is the key to the subsample results because it leads to the basic result given in Lemma 6(e).

Let $C(J_g)$ denote the set of continuity points of J_h.

Lemma 6 *Suppose Assumptions* B2 *and* S *hold. Let* $\{\gamma_{n,h} \in \Gamma : n \geq 1\}$ *satisfy* (i) $n^r \gamma_{n,h} \to h$ *and* (ii) $b_n^r \gamma_{n,h} \to g$ *for some* $g \in [0, \infty]$. *Then,*

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

(b) $E_{\theta_0, \gamma_{n,h}} U_{n,b_n}(x) = P_{\gamma_{n,h}}(T_{b_n}(\theta_0) \leq x) \to J_g(x)$ *for all* $x \in C(J_g)$,

(c) $U_{n,b_n}(x) \to_p J_g(x)$ *for all* $x \in C(J_g)$ *under* $\{\gamma_{n,h} : n \geq 1\}$,

(d) $c_{n,b_n}(1-\alpha) \to_p c_g(1-\alpha)$ *under* $\{\gamma_{n,h} : n \geq 1\}$,

(e) $P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) \leq c_{n,b_n}(1-\alpha)) \to J_h(c_g(1-\alpha))$, *and*

(f) *parts* (a)-(e) *of the Lemma hold with the sequence* $\{n\}$ *replaced by any subsequence* $\{w_n : n \geq 1\}$ *throughout the assumptions and results.*

Proof of Lemma 6. Part (a) holds because if $h < \infty$ and $b_n/n \to 0$, then $b_n^r \gamma_{n,h} \to 0$, $g = 0$, and $(g, h) \in GH$. On the other hand, if $h = \infty$, then $(g, h) \in GH$ because the definition of GH puts no restriction on g.

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

$$E_{\gamma_{n,h}} U_{n,b_n}(x) = q_n^{-1} \sum_{j=1}^{q_n} P_{\gamma_{n,h}}(T_{n,b_n,j}(\theta_0) \leq x) = P_{\gamma_{n,h}}(T_{b_n}(\theta_0) \leq x), \tag{11.30}$$

where the first equality holds by definition of $U_{n,b_n}(x)$ and the second equality holds by Assumption S(iii).

We now show that $P_{\gamma_{n,h}}(T_{b_n}(\theta_0) \leq x) \to J_g(x)$ for all $x \in C(J_g)$ by showing that any subsequence $\{t_n\}$ of $\{n\}$ has a sub-subsequence $\{s_n\}$ for which $P_{\gamma_{s_n,h}}(T_{b_{s_n}}(\theta_0) \leq x) \to J_g(x)$. Given any subsequence $\{t_n\}$, select a sub-subsequence $\{s_n\}$ such that $\{b_{s_n}\}$ is strictly increasing. This can be done because $b_n \to \infty$ by Assumption S(i). Because $\{b_{s_n}\}$ is strictly increasing, it is a subsequence of $\{n\}$. Below we show that Assumption B2 implies that for any subsequence $\{u_n\}$ of $\{n\}$ and any sequence $\{\gamma^*_{u_n} \in \Gamma : n \geq 1\}$ that satisfies (i′) $u_n^r \gamma^*_{u_n} \to g$, we have

$$P_{\gamma^*_{u_n}}(T_{u_n}(\theta_0) \leq x) \to J_g(x), \tag{11.31}$$

for all $x \in C(J_g)$. We apply this result with $u_n = b_{s_n}$ and $\gamma^*_{u_n} = \gamma_{s_n,h}$ to obtain the desired result $P_{\gamma_{s_n,h}}(T_{b_{s_n}}(\theta_0) \leq x) \to J_g(x)$, where (i′) holds because $b_{s_n}^r \gamma_{s_n,h} \to g$ by assumption.

For the proof of part (b), it remains to show (11.31). Define a new sequence $\{\gamma_k^{**} : k \geq 1\}$ as follows. If $k = u_n$ set γ_k^{**} equal to $\gamma_{u_n}^*$. If $k \neq u_n$, define γ_k^{**} to be

$$\begin{array}{ll} \gamma_k^{**} = \min\{k^{-2r}, b/2\} & \text{if } g = 0 \\ \gamma_k^{**} = \min\{k^{-r}g, b/2\} & \text{if } 0 < g < \infty \\ \gamma_k^{**} = b/2 & \text{if } g = \infty. \end{array} \tag{11.32}$$

Note that the parameters $\{\gamma_k^{**} : k \geq 1\}$ are in Γ for all $k \geq 1$ and $k^r \gamma_k^{**} \to g$ as $k \to \infty$. Hence, $\{\gamma_k^{**} : k \geq 1\}$ is of the form $\{\gamma_{n,g} : n \geq 1\}$ and Assumption B2 implies that $P_{\gamma_k^{**}}(T_k(\theta_0) \leq y) \to J_g(y)$ 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 $P_{\gamma_{u_n}^*}(T_{u_n}(\theta_0) \leq y) \to J_g(y)$, as desired.

Part (c) holds by part (b) and Assumption S(iv). Parts (d) and (e) are established by fairly standard arguments using part (c) and Assumption S(v) because $c_{n,b_n}(1-\alpha)$ is the $1-\alpha$ quantile of $U_{n,b_n}(x) = L_{n,b_n}(x)$ using Assumption S(vi). In particular, Lemma 3 in Appendix A can be used for this purpose. Part (f) holds by variations of the arguments given above with w_n in place of n throughout, see the proof of Lemma 4 for details. $\square$

Lemma 7 *Suppose Assumptions* B2 *and* S *hold. Let $(g,h) \in GH$ be given. Then, there is a sequence $\{\gamma_{n,h} : n \geq 1\}$ of points in Γ that satisfies $n^r \gamma_{n,h} \to h$ and $b_n^r \gamma_{n,h} \to g$ and for this sequence parts* (b)-(e) *of Lemma* 6 *hold.*

Proof of Lemma 7. Define a sequence $\{\gamma_{n,h} : n \geq 1\}$ as follows:

$$\begin{array}{ll} \gamma_{n,h} = \min\{n^{-2r}, b/2\} & \text{if } g = h = 0 \\ \gamma_{n,h} = \min\{n^{-r}h, b/2\} & \text{if } g = 0 \ \& \ 0 < h < \infty \\ \gamma_{n,h} = \min\{(b_n n)^{-r/2}, b/2\} & \text{if } g = 0 \ \& \ h = \infty \\ \gamma_{n,h} = \min\{b_n^{-r}g, b/2\} & \text{if } 0 < g < \infty \ \& \ h = \infty \\ \gamma_{n,h} = b/2 & \text{if } g = h = \infty. \end{array} \tag{11.33}$$

As defined, $\gamma_{n,h} \in (0, b] \subset \Gamma$ for all $n \geq 1$ and straightforward calculations show that $\{\gamma_{n,h} : n \geq 1\}$ satisfies $n^r \gamma_{n,h} \to h$ and $b_n^r \gamma_{n,h} \to g$. $\square$

Proof of Corollary 1. The proof of part (a) is similar to that of part (b), but much simpler. Hence, for brevity, we only give the proof for part (b). We first show that $AsySz(\theta_0) \geq Max_{Sub}(\alpha)$. Equations (3.10) and (3.11) imply that for any sequence $\{\gamma_n \in \Gamma : n \geq 1\}$,

$$AsySz(\theta_0) \geq \limsup_{n\to\infty} [1 - P_{\theta_0,\gamma_n}(T_n(\theta_0) \leq c_{n,b}(1-\alpha))]. \tag{11.34}$$

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 $\{\gamma_n \in \Gamma : n \geq 1\}$ such that

$$\limsup_{n\to\infty} [1 - P_{\theta_0,\gamma_n}(T_n(\theta_0) \leq c_{n,b}(1-\alpha))] \geq 1 - J_h(c_g(1-\alpha)). \tag{11.35}$$

The latter inequality holds (as an equality) by Lemma 7.

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

Either (1) $\limsup_{n\to\infty} v_n^r \gamma_{v_n}^* < \infty$ or (2) $\limsup_{n\to\infty} v_n^r \gamma_{v_n}^* = \infty$. If (1) holds, then for some subsequence $\{w_n\}$ of $\{v_n\}$

$$\begin{aligned} w_n^r \gamma_{w_n}^* &\to h \text{ for some } h \in R \text{ and} \\ b_{w_n}^r \gamma_{w_n}^* &\to 0. \end{aligned} \tag{11.36}$$

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

$$\begin{aligned} b_{w_n}^r \gamma_{w_n}^* &\to g \text{ for some } g \in R \text{ and} \\ w_n^r \gamma_{w_n}^* &\to h, \text{ where } h = \infty. \end{aligned} \tag{11.37}$$

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

$$\begin{aligned} b_{w_n}^r \gamma_{w_n}^* &\to g, \text{ where } g = \infty, \text{ and} \\ w_n^r \gamma_{w_n}^* &\to h, \text{ where } h = \infty. \end{aligned} \tag{11.38}$$

In consequence, conditions (i) and (ii) of Lemma 6 hold. Hence,

$$RP_{w_n}(\theta_0, \gamma_{w_n}^*) \to 1 - J_h(c_g(1-\alpha)) \tag{11.39}$$

by Lemma 6(e) with w_n in place of n, which holds by Lemma 6(f). Also, $(g, h) \in GH$ by Lemma 6(a). Since $\lim_{n\to\infty} RP_{v_n}(\theta_0, \gamma_{v_n}^*) = AsySz(\theta_0)$ and $\{w_n\}$ is a subsequence of $\{v_n\}$, we have $\lim_{n\to\infty} RP_{w_n}(\theta_0, \gamma_{w_n}^*) = AsySz(\theta_0)$. This, (11.39) and $(g, h) \in GH$ imply that $AsySz(\theta_0) \leq Max_{Sub}(\alpha)$, which completes the proof of part (b). □

11.2 CI for the Endpoint of a Distribution (Cont.)

Here we consider an equal-tailed subsample CI for the lower endpoint, θ, of the support of a distribution. We show that this CI with nominal level $1-\alpha$ has $AsyCS = 1 - \alpha$, as desired. The subsample CI has correct asymptotic size because this is a "continuous limit" example as a function of the parameters in the model (and the limit distribution function is continuous at the appropriate quantiles).

The observations $\{X_i : i = 1, ..., n\}$ are i.i.d. with distribution F, where F has a density f with respect to Lebesgue measure and support with finite lower endpoint $\theta = F^{-1}(0)$. For example, F could be $U[\theta, 1]$. This example is non-regular in that rate-n consistent estimators of θ are available.

Because the present example is a "continuous limit" example, there is no parameter γ_1. We define $\gamma_2 = (\theta, f(\theta))'$ and $\gamma_3 = F$. The parameter space for γ_2 is $\Gamma_2 = \{(\theta, \gamma_{22})' :$

$\theta \in R, B_1 \leq \gamma_{22} \leq B_2\}$ for some $0 < B_1 < B_2 < \infty$. The parameter space for γ_3 is $\Gamma_3(\gamma_2) = \{F : F$ has a density f with respect to Lebesgue measure that satisfies (i) $f(\theta) = \gamma_{22}$, where $\gamma_2 = (\theta, \gamma_{22})'$, and (ii) f is right continuous at θ with modulus of continuity $\delta \leq M\}$, for some $M < \infty$. The definition of Γ_2 bounds the density of F away from zero at θ. This ensures that rate-n consistent estimators of θ are available. With some added complication, this assumption could be relaxed, see Loh (1984).

We consider the maximum likelihood estimator of θ: $\widehat{\theta}_n = X_{(1)}$, where $X_{(j)}$ denotes the jth order statistic from $\{X_i : i = 1, ..., n\}$. The test statistic is

$$T_n(\theta) = n(X_{(1)} - \theta). \tag{11.40}$$

In this example, $\tau_n = n$. No rate of convergence parameter r arises in this continuous limit example because there is no parameter γ_1.

The equal-tailed subsample CI of θ is defined as in (8.5) with $\widehat{\sigma}_n = 1$, $c_{1-\alpha/2} = c_{n,b}(1 - \alpha/2)$, and $c_{\alpha/2} = c_{n,b}(\alpha/2)$.

We have $H = H_2 = cl(\Gamma_2)$. We consider sequences $\{\gamma_{n,h} = (\gamma_{2,n,h}, \gamma_{3,n,h}) = ((\theta_{n,h}, f_{n,h}(\theta_{n,h}))', F_{n,h}) : n \geq 1\}$ that are defined such that $\gamma_{2,n,h} = (\theta_{n,h}, f_{n,h}(\theta_{n,h}))' \in \Gamma_2$, $\gamma_{2,n,h} \to h_2$ for some $h_2 = (h_{21}, h_{22})' \in H_2$, and $\gamma_{3,n,h} \in \Gamma_3(\gamma_{2,n,h})$. As defined, the parameter h_{22} is the limit as $n \to \infty$ of $f_{n,h}(\theta_{n,h})$. For any such sequence $\{\gamma_{n,h} : n \geq 1\}$, we have

$$\begin{aligned} &T_n(\theta_{n,h}) \to_d J_h, \text{ where} \\ &J_h(x) = 1 - \exp\{-h_{22}x\} \text{ for } x \geq 0. \end{aligned} \tag{11.41}$$

That is, J_h is an exponential distribution with parameter $h_{22} \in [B_1, B_2]$. Hence, Assumption B2 holds. The proof is given below.

It is straightforward to verify that Assumptions A2, C, D, E, t1, t2, Sub1, and H hold. Assumption J2 holds because J_h has support R_+ and is strictly increasing on R_+. Assumption G2 holds by Lemma 1. By Corollary 4(b),

$$AsyCS \in [1 - Max^{\ell-}_{ET,Sub}(\alpha), 1 - Max^{r-}_{ET,Sub}(\alpha)], \tag{11.42}$$

where $Max^{\ell-}_{ET,Sub}(\alpha)$ and $Max^{r-}_{ET,Sub}(\alpha)$ are defined in (7.3). Because J_h is continuous at its τ quantile for all $\tau \in (0, 1)$, we have $Max^{\ell-}_{ET,Sub}(\alpha) = Max^{r-}_{ET,Sub}(\alpha)$. Furthermore, because no parameter γ_1 appears, the parameter space GH reduces to $\{(g_2, h_2) \in H_2 \times H_2 : g_2 = h_2\}$ and $Max^{\ell-}_{ET,Sub}(\alpha)$ simplifies to

$$Max^{\ell-}_{ET,Sub}(\alpha) = \sup_{h_2 \in H_2} [1 - J_{h_2}(c_{h_2}(1 - \alpha/2)) + J_{h_2}(c_{h_2}(\alpha/2))] = \alpha. \tag{11.43}$$

We conclude that the equal-tailed subsample CI has $AsyCS = 1 - \alpha$. Analogous results show that the subsample CI is asymptotically similar even though the asymptotic distribution of $T_n(\theta)$ is nuisance parameter dependent.

It remains to verify Assumption B2 with J_h as in (11.41). For $x \geq 0$, we have

$$P_{F_{n,h}}(T_n(\theta_{n,h}) > x) = P_{F_{n,h}}(X_i > x/n + \theta_{n,h}\ \forall i \leq n) = (1 - F_{n,h}(x/n + \theta_{n,h}))^n. \tag{11.44}$$

By a mean-value expansion, we obtain

$$\begin{aligned} F_{n,h}(x/n + \theta_{n,h}) &= F_{n,h}(\theta_{n,h}) + f_{n,h}(\theta^*_{n,h})x/n \\ &= f_{n,h}(\theta^*_{n,h})x/n \\ &= [f_{n,h}(\theta_{n,h}) + o(1)]x/n, \end{aligned} \tag{11.45}$$

where $\theta^*_{n,h}$ lies between $\theta_{n,h}$ and $\theta_{n,h} + x/n$, $F_{n,h}(\theta_{n,h}) = 0$ because $\theta_{n,h}$ is the lower endpoint of the support of $F_{n,h}$, and the third equality holds because $f_{n,h}$ is right continuous at $\theta_{n,h}$ with modulus of continuity $\delta \le M$ by the definition of $\Gamma_3(\gamma_2)$. Substitution of (11.45) into (11.44) gives $P_{F_{n,h}}(T_n(\theta_{n,h}) > x) \to \exp\{-h_{22}x\}$ as desired.

References

Anatolyev, S. (2004): "Inference When a Nuisance Parameter Is Weakly Identified under the Null Hypothesis," *Economics Letters,* 84, 245-254.

Andrews, D. W. K. (1993): "Exactly Median-Unbiased Estimation of First-Order Autoregressive/Unit Root Models," *Econometrica,* 61, 139-165.

——— (1999): "Estimation When a Parameter Is on a Boundary," *Econometrica,* 67, 1341–1383.

——— (2000): "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space," *Econometrica,* 68, 399–405.

——— (2001): "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica,* 69, 683–734.

Andrews, D. W. K. and P. Guggenberger (2005a): "Hybrid and Size-corrected Subsample Methods," unpublished manuscript, Cowles Foundation, Yale University.

——— (2005b): "Applications of Hybrid and Size-Corrected Subsample Methods," unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. and J. H. Stock (2006): "Inference with Weak Instruments," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society,* Vol. III, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge, UK: Cambridge University Press, forthcoming. Also available as Cowles Foundation Discussion Paper No. 1530, 2005.

Babu, G. J. (1984): "Bootstrapping Statistics with Linear Combinations of Chi-Squares as Weak Limits," *Sankhya,* 46, 85-93.

Bahadur, R. R. and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Annals of Mathematical Statistics,* 25, 1115-1122.

Beran, R. (1997): "Diagnosing Bootstrap Success," *Annals of the Institute of Statistical Mathematics,* 49, 1-24.

Beran, R. and M. S. Srivastava (1985): "Bootstrap Tests and Confidence Regions for Functions of a Covariance Matrix," *Annals of Statistics,* 13, 95-115. Correction, 1987, *Annals of Statistics,* 15, 470-471.

——— (1987): "Correction: Bootstrap Tests and Confidence Regions for Functions of a Covariance Matrix," *Annals of Statistics,* 15, 470-471.

Bickel, P. J. and 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," *Annales de l'Institut Henri Poincaré,* Sec. B, 19, 281-296.

Cavanagh, C. L., G. Elliot, and J. H. Stock (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory,* 11, 1131-1147.

Chernozhukov, V. and H. Hong (2004): "Likelihood Estimation and Inference in Nonregular Models," *Econometrica,* 72, 1445-1480.

Chernozhukov, V., H. Hong, and E. Tamer (2002): "Parameter Set Inference in a Class of Econometric Models," unpublished working paper, Department of Economics, MIT.

Choi, I. (2005): "Subsampling Vector Autoregressive Tests of Linear Constraints," *Journal of Econometrics,* 124, 55-89.

Dufour, J.-M. (1997): "Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica,* 65, 1365-1387.

Dümbgen, L. (1993): "On Nondifferentiable Functions and the Bootstrap," *Probability Theory and Related Fields,* 95, 125-140.

Eaton, M. L. and 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.

Efron, B. and R. J. Tibshirani (1993): *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Galichon, A. and M. Henry (2006): "Inference in Incomplete Models," unpublished manuscript, Department of Economics, Columbia University.

Gleser, L. J. and J. T. Hwang (1987): "The Nonexistence of 100(1-α)% Confidence Sets of Finite Expected Diameter in Errors-in-Variables and Related Models," *Annals of Statistics,* 15, 1351-1362.

Guggenberger, P. and M. Wolf (2004): "Subsampling Tests of Parameter Hypotheses and Overidentifying Restrictions with Possible Failure of Identification," unpublished working paper, Department of Economics, UCLA.

Hájek, J. (1971): "Limiting Properties of Likelihoods and Inference," in *Foundations of Statistical Inference: Proceedings of the Symposium on the Foundations of Statistical Inference, University of Waterloo, Ontario,* ed. by V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart, and Winston, 142-159.

Hall, P. and B. Jing (1995): "Uniform Coverage Bounds for Confidence Intervals and Berry-Esseen Theorems for Edgeworth Expansion," *Annals of Statistics,* 23, 363-375.

Hirano, K. and J. R. Porter (2003): "Asymptotic Efficiency in Parametric Models with Parameter-Dependent Support," *Econometrica,* 71, 1307-1338.

Imbens, G. and 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.

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. and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory,* 21, 21-59.

——— (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.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies,* 72, 735-765.

Loh, W.-Y. (1984): "Estimating an Endpoint of a Distribution with Resampling Methods," *Annals of Statistics,* 12, 1543-1550.

——— (1985): "A New Method for Testing Separate Families of Hypotheses," *Journal of the American Statistical Association,* 80, 362-368.

Mikusheva, A. (2005): "Uniform Inferences in Autoregressive Models," unpublished working paper, Department of Economics, Harvard University.

Moon, H. R. and F. Schorfheide (2004): "Empirical Likelihood Estimation with Inequality Moment Conditions," unpublished working paper, Department of Economics, University of Southern California.

Pfanzagl, J. (1973): "The Accuracy of the Normal Approximation for Estimates of Vector Parameters," *Z. Wahrscheinlichkeitstheorie verw. Geb.,* 25, 171-198.

Politis, D. N. and 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, and M. Wolf (1999): *Subsampling.* New York: 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.

——— (1973): *Linear Statistical Inference and Its Applications.* Second edition. New York: Wiley.

Romano, J. P. (1989): "Do Bootstrap Confidence Procedures Behave Well Uniformly in P?," *Canadian Journal of Statistics,* 17, 75-80.

Romano, J. P. and A. M. Shaikh (2005a): "Inference for Identifiable Parameters in Partially Identified Econometric Models," unpublished working paper, Department of Economics, University of Chicago.

——— (2005b): "Inference for the Identified Set in Partially Identified Econometric Models," unpublished working paper, Department of Economics, University of Chicago.

Romano, J. P. and M. Wolf (2001): "Subsampling Intervals in Autoregressive Models with Linear Time Trend," *Econometrica,* 69, 1283–1314.

Rosen, A. M. (2005): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," unpublished working paper, Department of Economics, Northwestern University.

Sen, P. K. (1979): "Asymptotic Properties of Maximum Likelihood Estimators Based on Conditional Specification," *Annals of Statistics, 7,* 1019-1033.

Sen, P. K. and 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.

——— (1996): "Bootstrap Model Selection," *Journal of the American Statistical Association,* 91, 655-665.

Shao, J. and C. J. F. Wu (1989): "A General Theory for Jackknife Variance Estimation," *Annals of Statistics,* 15, 1563-1579.

Sherman, M. and E. Carlstein (1996): "Replicate Histograms," *Journal of the American Statistical Association,* 91, 566-576.

Soares, G. (2005): "Inference with Inequality Moment Constraints," unpublished working paper, Department of Economics, Yale University.

Staiger, D. and 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.

Swanepoel, J. W. H. (1986): "A Note on Proving That the (Modified) Bootstrap Works," *Communications in Statistics—Theory and Methods,* 15, 3193-3203.

Wu, C. F. J. (1990): "On the Asymptotic Properties of the Jackknife Histogram," *Annals of Statistics,* 18, 1438-1452.

TABLE I

NUISANCE PARAMETER NEAR A BOUNDARY EXAMPLE: MAXIMUM ASYMPTOTIC NULL REJECTION PROBABILITIES (×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	9.9	5.0	52.7	5.0
-.99	42.8	5.0	9.9	5.0	43.2	5.0
-.95	33.8	5.0	9.9	5.0	32.4	5.0
-.90	27.6	5.0	9.9	5.0	25.4	5.0
-.80	20.2	5.0	9.3	5.0	17.4	5.0
-.60	12.3	5.0	7.4	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	5.3	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	6.7	5.0
.60	5.0	5.6	7.5	5.0	9.9	5.0
.80	5.0	5.1	9.6	5.0	17.3	5.0
.90	5.0	5.0	10.1	5.0	25.2	5.0
.95	5.0	5.0	10.1	5.0	32.4	5.0
.99	5.0	5.0	10.1	5.0	43.0	5.0
1.00	5.0	5.0	10.1	5.0	52.3	5.0
$AsySz(\theta_0)$	50.2	5.8	10.1	5.0	52.7	5.0

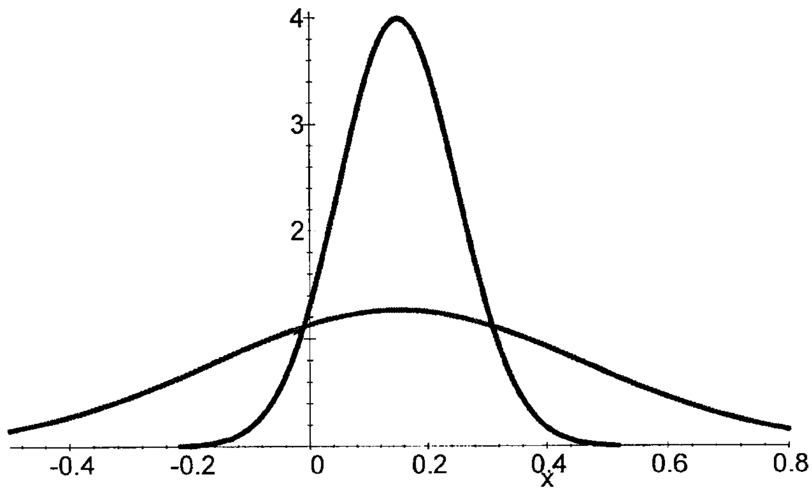

FIGURE 1.—Densities of $Z_n \sim N(\theta_0, n^{-1})$ (peaked curve) and $Z_{b_n} \sim N(\theta_0, b_n^{-1})$ (flatter curve) for $\theta_0 = .15$, $n = 100$, and $b_n = 10$

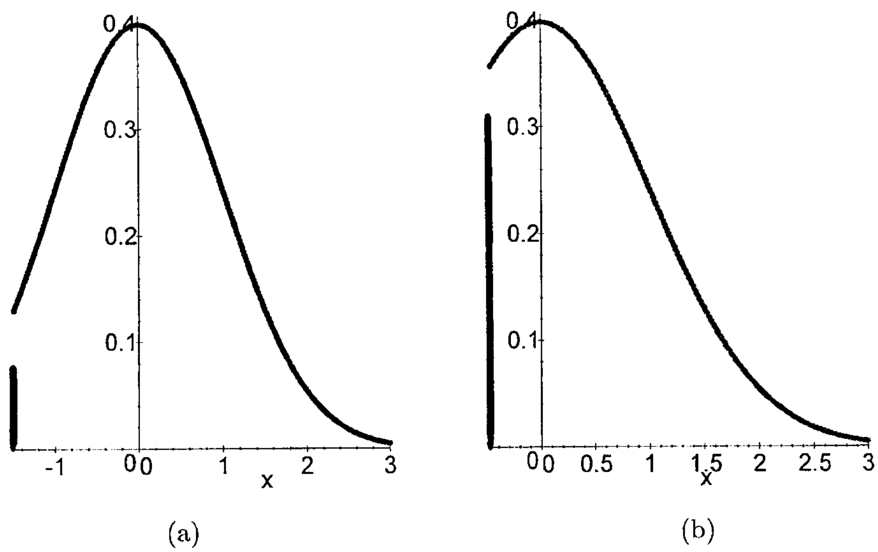

FIGURE 2.—Densities of (a) $T_n \sim \max\{Z, -h\}$ and (b) $T_{b_n} \sim \max\{Z, -(b_n/n)^{1/2}h\}$ for $Z \sim N(0,1)$, $h = 1.5$, and $b_n/n = 1/10$

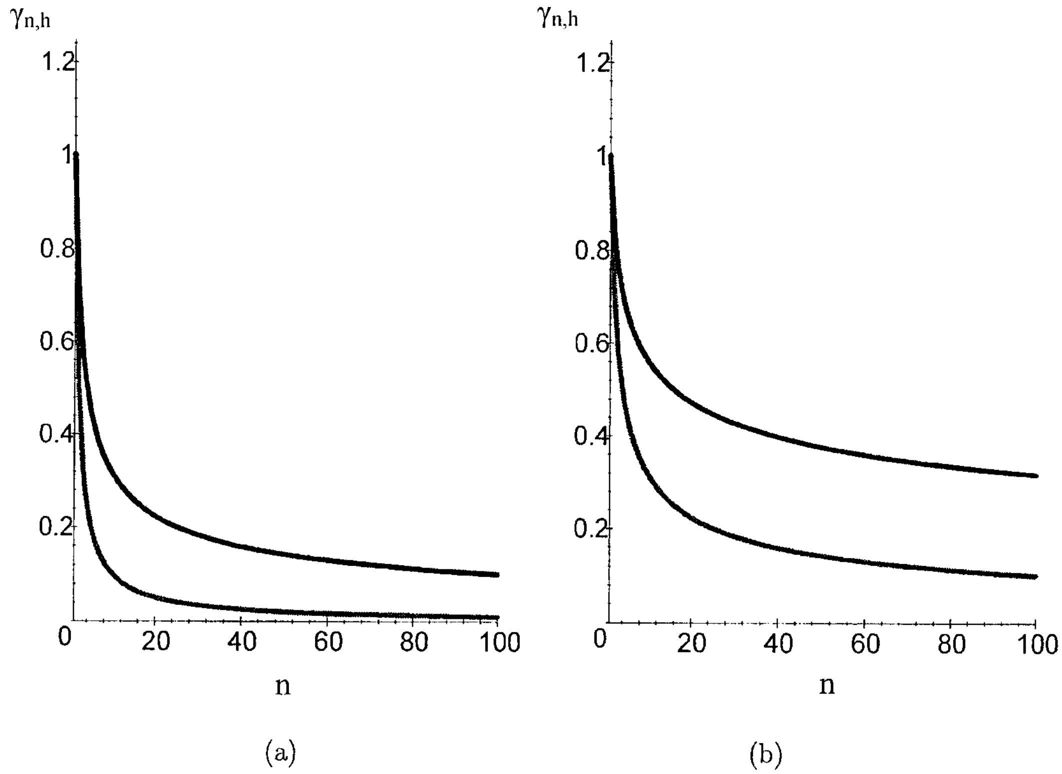

FIGURE 3.—$\gamma_{n,h}$ as a function of n (upper curve) and as a function of $b_n = n^{1/2}$ (lower curve) for (a) $\gamma_{n,h} = 1/n^{1/2}$ and (b) $\gamma_{n,h} = 1/n^{1/4}$

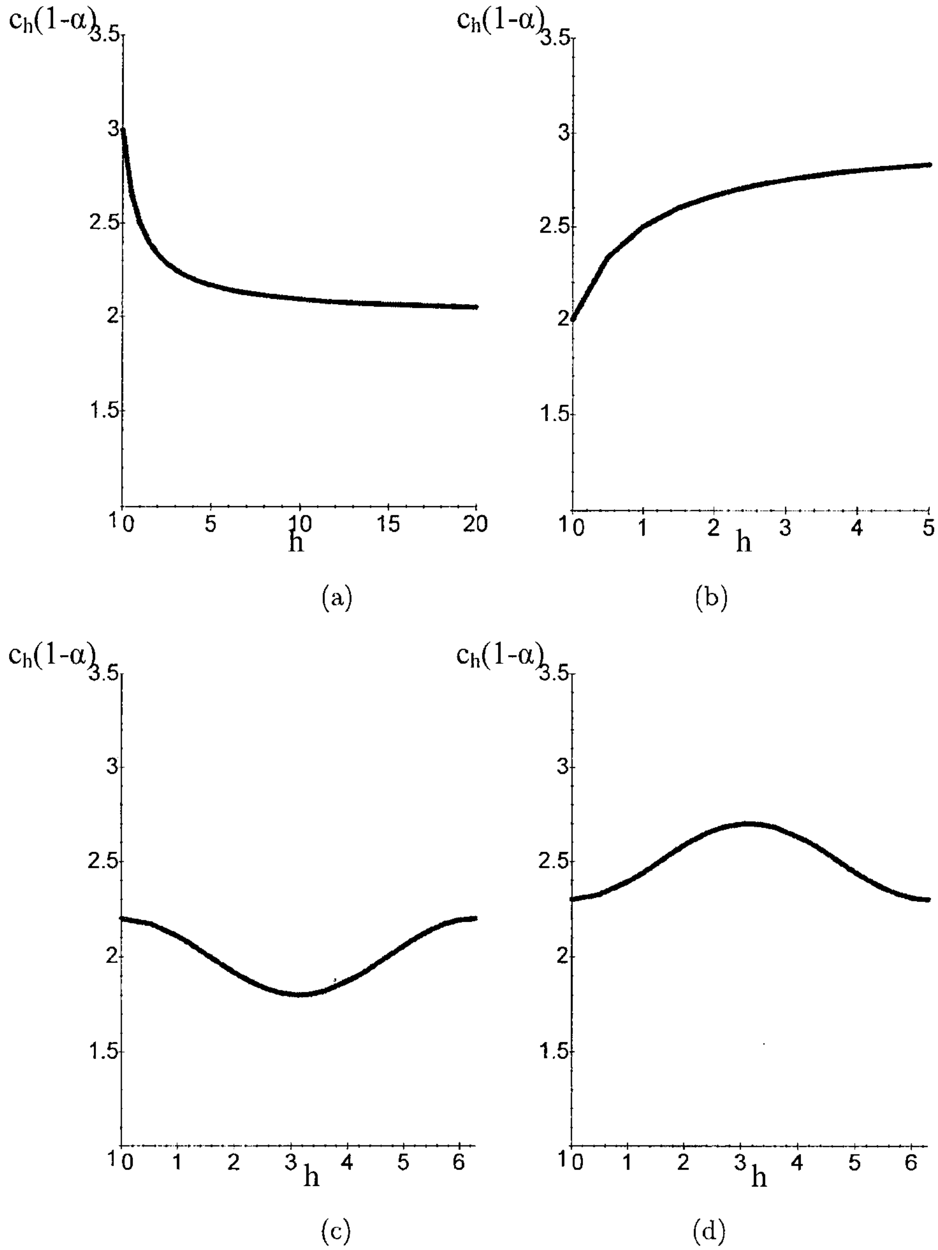

FIGURE 4.—$1-\alpha$ quantile of J_h, $c_h(1-\alpha)$, as a function of h

FIGURE 5.—Nuisance Parameter Near a Boundary Example: .95 Quantile Graphs, $c_h(.95)$, for J_h^* and $|J_h^*|$ as Functions of h_1 for Several Values of the Correlation h_2