INCORRECT ASYMPTOTIC SIZE OF SUBSAMPLING PROCEDURES BASED ON POST-CONSISTENT MODEL SELECTION ESTIMATORS

BY

DONALD W. K. ANDREWS and PATRIK GUGGENBERGER

Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators

Donald W.K. Andrews [a,*], Patrik Guggenberger [b]

[a] *Cowles Foundation for Research in Economics, Yale University, United States*

[b] *Department of Economics, UCLA, United States*

ABSTRACT

Subsampling and the m out of n bootstrap have been suggested in the literature as methods for carrying out inference based on post-model selection estimators and shrinkage estimators. In this paper we consider a subsampling confidence interval (CI) that is based on an estimator that can be viewed either as a post-model selection estimator that employs a consistent model selection procedure or as a super-efficient estimator. We show that the subsampling CI (of nominal level $1 - \alpha$ for any $\alpha \in (0, 1)$) has asymptotic confidence size (defined to be the limit of finite-sample size) equal to zero in a very simple regular model. The same result holds for the m out of n bootstrap provided $m^2/n \to 0$ and the observations are i.i.d. Similar zero-asymptotic-confidence-size results hold in more complicated models that are covered by the general results given in the paper and for super-efficient and shrinkage estimators that are not post-model selection estimators. Based on these results, subsampling and the m out of n bootstrap are not recommended for obtaining inference based on post-consistent model selection or shrinkage estimators.

© 2009 Published by Elsevier B.V.

1. Introduction

Over the years, Peter Robinson has made path breaking contributions to time series analysis. Particularly noteworthy is his contribution to a class of non-regular time series models, viz., those with long memory. In this paper, we consider inference in the presence of a different non-regular feature, viz., inference based on a test statistic that has a discontinuity in its asymptotic distribution as a function of some parameter. In particular, we consider subsampling inference based on post-model selection test statistics and test statistics based on shrinkage estimators. Post-model selection tests and confidence intervals (CIs) are widely used in practice, with both time series and cross-section observations. Shrinkage estimators are important because many new developments in nonparametric statistics rely on shrinkage methods. Subsampling a super-efficient estimator has been

suggested in Politis et al. (1999) (hereafter PRW) and Lehmann and Romano (2005). Using the m out of n bootstrap for a post-model selection estimator has been suggested by Shao (1994, 1996).

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 PRW.[1] Minimal conditions are needed for subsampling 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.

Recent papers by Andrews and Guggenberger (forthcoming-a,c,d) and Guggenberger (2008, Supplement), however, show that subsampling methods often do not yield the correct asymptotic size (defined to be the limit of finite-sample size) when a test

---

* Corresponding address: Department of Economics, Yale University, 30 Hillhouse Ave, Rm. 17, Box 208281, New Haven, CT 06520-8281, United States. Tel.: +1 203 432 3698; fax: +1 203 432 6167.

E-mail address: donald.andrews@yale.edu (D.W.K. Andrews).

---

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

statistic has a discontinuity in its asymptotic distribution.[2] The results in those papers rely on an assumption, viz., Assumption B, that is violated in the cases of post-model selection inference based on "consistent" model selection procedures and inference based on shrinkage estimators; see Section 3.3.[3] In this paper we provide results based on a weaker condition than Assumption B, viz., Assumption B1, and show that this assumption holds in the cases above.

The results of this paper establish that subsampling CIs based on a post-consistent model selection or shrinkage estimator have asymptotic size equal to zero in a very simple model. The general results provided in the paper also apply to more complicated models.

The reason why the subsampling CI has asymptotic size equal to zero in the present context is explained as follows. Consider a shrinkage estimator $\widehat{\theta}_n$ of a parameter θ that equals zero when some preliminary estimator $\widehat{\theta}_{P,n}$ satisfies $|\widehat{\theta}_{P,n}| \leq n^{-1/4}$ and otherwise equals $\widehat{\theta}_{P,n}$, where n is the sample size. A nominal $1 - \alpha$ subsampling CI for θ is obtained by inverting subsampling tests based on the test statistic $T_n(\theta_0) = |n^{1/2}(\widehat{\theta}_n - \theta_0)|$ for testing $H_0 : \theta = \theta_0$ for each $\theta_0 \in R$.[4] This CI is of the usual form $\widehat{\theta}_n \pm c_{n,b}(1 - \alpha)/n^{1/2}$, where $c_{n,b}(1 - \alpha)$ is the subsampling critical value based on subsamples of size b, where $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. The subsampling critical value, $c_{n,b}(1 - \alpha)$, is the $1 - \alpha$ sample quantile of the subsample statistics of size b. The shrinkage estimator $\widehat{\theta}_n$ can be viewed as a post-consistent model selection estimator where model 1 takes $\theta = 0$ and model 2 takes $\theta \in R$.

Suppose the true parameter is $\theta_{n,h} = h/n^{1/2}$ for some constant $h \neq 0$ for $n \geq 1$. We consider the asymptotic behavior of $T_n(\theta_{n,h})$ and $c_{n,b}(1 - \alpha)$ under $\{\theta_{n,h} : n \geq 1\}$. As is typical, suppose the preliminary estimator $\widehat{\theta}_{P,n}$ satisfies $n^{1/2}(\widehat{\theta}_{P,n} - \theta_{n,h}) = O_p(1)$ and $b^{1/2}(\widehat{\theta}_{P,b} - \theta_{n,h}) = O_p(1)$ under $\{\theta_{n,h} : n \geq 1\}$. Under $\{\theta_{n,h} : n \geq 1\}$, $\widehat{\theta}_n = 0$ with probability that goes to one (wp $\to$ 1), because $|\widehat{\theta}_{P,n}| = |\widehat{\theta}_{P,n} - \theta_{n,h}| + O_p(n^{-1/2}) = O_p(n^{-1/2}) < n^{-1/4}$wp $\to$ 1. In consequence, $T_n(\theta_{n,h}) = |n^{1/2}\theta_{n,h}| = |h|$ wp$\to$ 1.

On the other hand, it can be shown that the asymptotic distribution of the subsample statistics of size b is the same as the asymptotic distribution of the full-sample statistic based on a full sample of size b, i.e., $T_b(\theta_{n,h})$. Hence, the probability limit of the

subsampling critical value is the $1 - \alpha$ quantile of the asymptotic distribution of $T_b(\theta_{n,h})$ under $\{\theta_{n,h} : n \geq 1\}$. By definition, $T_b(\theta_{n,h}) = |b^{1/2}(\widehat{\theta}_b - \theta_{n,h})|$ and $\widehat{\theta}_b = 0$ if $|\widehat{\theta}_b| \leq b^{-1/4}$. The latter occurs wp $\to$ 1 because $|\widehat{\theta}_{P,b}| = |\widehat{\theta}_{P,b} - \theta_{n,h}| + O_p(n^{-1/2}) = O_p(b^{-1/2}) < b^{-1/4}$ wp$\to$ 1. Hence, $\widehat{\theta}_b = 0$ wp $\to$ 1 and $T_b(\theta_{n,h}) = |b^{1/2}\theta_{n,h}| + o_p(1) = o_p(1)$ since $b/n \to 0$. In turn, this implies that the subsampling critical value converges in probability to 0. Since the test statistic $T_n(\theta_{n,h})$ converges in probability to $|h| > 0$ and the subsampling critical value converges in probability to 0, the subsampling test of $H_0 : \theta = \theta_{n,h}$ rejects wp $\to$ 1 and the CI obtained by inverting the subsampling tests fails to include the true value $\theta_{n,h}$ wp$\to$ 1. This implies that the finite-sample confidence size of the subsampling CI goes to zero as $n \to \infty$.

In short, the argument above is: $T_n(\theta_{n,h}) \to_p |h|$ under $\theta_{n,h} = (h + o(1))/n^{1/2}$ for any $h \in R$ implies that $T_b(\theta_{b,h}) \to_p |h|$ under $\theta_{b,h} = (h + o(1))/b^{1/2}$ for all $h \in R$, and so, $T_b(\theta_{n,h}) \to_p 0$ under $\theta_{n,h} = (h + o(1))/n^{1/2} = (0 + o(1))/b^{1/2}$ since $b/n \to 0$. So, the subsample statistics are smaller than the full-sample statistic in large samples and the subsampling test rejects wp $\to$ 1.

The paper also gives results for CIs that are based on fixed critical values (FCVs) rather than subsampling critical values. The asymptotic results given here for subsampling tests also apply to m out of n bootstrap tests applied to i.i.d. observations provided $m^2/n \to 0$. The reason is that subsampling based on subsamples of size m can be viewed as bootstrapping without replacement, which is not too different from bootstrapping with replacement when m^2/n is small.[5]

Related results in the literature include the following. Samworth (2003) provides simulation results and heuristics indicating that the m out of n bootstrap does not provide a good approximation to the distribution of a shrinkage estimator.[6] Beran (1982) shows that the (standard) bootstrap is inconsistent for the distribution of a shrinkage estimator. Kabaila (1995) shows that an FCV CI based on a super-efficient estimator has asymptotic confidence level equal to zero; see Leeb and Pötscher (2005) for related results. The results of Leeb and Pötscher (2006) show that no uniformly consistent estimator of the distribution of a super-efficient estimator exists. The results given in this paper 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. For example, Andrews and Guggenberger (forthcoming-b) provides an example which illustrates this in the context of inference based on moment inequalities.

Subsequent to the present paper, Pötscher (unpublished manuscript) has provided some results concerning confidence sets based on sparse estimators. Some, but not all, of the subsampling results of the present paper also can be established via results in Pötscher (unpublished manuscript). In particular, Pötscher's (unpublished manuscript) results do not provide an expression for the asymptotic coverage probability of a subsampling CI as a function of the localization parameter h, which is the main result of this paper; see Theorem 2 below. In addition, Pötscher's (unpublished manuscript) results do not apply to (i) the specific example considered below with $a > 0$ because the estimator is not a sparse estimator and (ii) models in which the subsampling

---

[2] The finite-sample (or exact) *size* of a test is defined to be the maximum rejection probability of the test under distributions in the null hypothesis. A test is said to have *level* α if its finite-sample size is α or less. The *asymptotic size* of a test is defined to be the limit of the finite-sample size of the test. The finite-sample (or exact) *size* of a confidence interval (or confidence set) is defined to be the minimum coverage probability of the confidence interval under distributions in the model. Analogously, a confidence interval is said to have *level* $1 - \alpha$ if its finite-sample size is $1 - \alpha$ or greater. The *asymptotic size* of a confidence interval is defined to be the limit of the finite-sample size of the confidence interval.

[3] A "consistent" model selection procedure, like BIC, selects the most parsimonious correct model with probability that goes to one as $n \to \infty$ under a fixed probability distribution for the data. A "conservative" model selection procedure, like AIC, selects a correct model, but not necessarily the most parsimonious correct model, with probability that goes to one as $n \to \infty$ under a fixed probability distribution for the data.

[4] Here we consider a non-studentized test statistic because subsampling tests do not require studentization given that changes in the scale of the test statistic cancel with the corresponding changes in the scale of the subsample statistics. For example, if the test statistic $T_n(\theta_0)$ is multiplied by $\tau > 0$, then by their definition the subsample statistics also are multiplied by τ, and the subsampling critical value is multiplied by τ, which leaves the test unchanged. This does not mean that a subsampling test is the same whether or not one studentizes the test statistic. It means that studentization is not necessary for subsampling tests to work properly when the purpose of studentization is to account for unknown scale τ. In other contexts, such as with unit roots, studentization may be necessary, but in those cases, studentization is doing more than just accounting for unknown scale τ.

Analogous results to those discussed here also hold for studentized statistics, as is shown in the general results below.

[5] 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 $m = b$ conditional on there being no duplicates of observations in the bootstrap sample. If $m^2/n \to 0$, then the probability of no duplicates goes to one as $n \to \infty$; see PRW, p. 48. In consequence, m out of n bootstrap tests and subsampling tests have the same first-order asymptotic properties.

[6] The simulation results in Samworth (2003) are for the case where the constant a, defined in (2.4) below, equals .5. For smaller values of a, such as $a = 0$, the results are exacerbated.

critical value $c_{n,b}(1 - \alpha)$ is not stochastically bounded uniformly in the parameter θ, which needs to be established for Pötscher's (unpublished manuscript) results to apply. For an example in which the latter condition fails, see Andrews and Guggenberger (forthcoming-d). This example concerns subsampling in a linear instrumental variables regression model with possibly weak instruments.

Other papers that consider uniformity properties of subsampling methods include (i) Andrews and Guggenberger (forthcoming-a,b,c,d), who provide explicit expressions for asymptotic size, improvements to subsampling based on hybrid and size-correction methods, and applications to a variety of different models, (ii) Mikusheva (2007), who shows that equal-tailed two-sided subsampling CIs do not have correct asymptotic size in an autoregressive model with a root that may be near unity, and (iii) Romano and Shaikh (2008), who provide high-level conditions under which subsampling CIs have correct asymptotic size and apply them to parameters defined by moment inequalities.

The remainder of the paper is organized as follows. Section 2 defines the class of FCV and subsampling CIs that are considered in the paper and introduces the post-consistent model selection/shrinkage estimator example. Section 3 states the general assumptions and verifies them in the post-consistent model selection/shrinkage estimator example. Section 4 states the general asymptotic results and shows that they imply that the post-consistent model selection/shrinkage estimator CI has asymptotic size equal to zero. Section 5 provides proofs of the general results.

2. Confidence interval set-up

2.1. Test statistics

We are interested in confidence intervals (CIs) (or confidence regions) for a parameter $\theta \in R^d$ in the presence of nuisance parameters. We construct such intervals by inverting a test statistic $T_n(\theta_0)$ for testing $H_0 : \theta = \theta_0$. The test statistic $T_n(\theta_0)$ may be an LR, LM, Wald, t, or some other statistic. 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, $\tau_n = n^{1/2}$.

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

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

We employ either a fixed critical value (FCV), $c_{Fix}(1 - \alpha)$, or a subsampling critical value, $c_{n,b}(1 - \alpha)$, defined below. Let $\Theta(\subset R^d)$ denote the parameter space for θ. 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:

$$CI_n = \{\theta_0 \in \Theta : T_n(\theta_0) \leq c_{1-\alpha}\}, \tag{2.1}$$

where $c_{1-\alpha}$ equals $c_{Fix}(1 - \alpha)$ or $c_{n,b}(1 - \alpha)$ (and $c_{1-\alpha}$ may depend on θ_0).

For example, suppose $T_n(\theta_0)$ is (i) an 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) and $c_{1-\alpha}$ does not depend on θ_0. Then, the corresponding CI of nominal level α is defined by

$$CI_n = [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}, \infty),$$
$$CI_n = (-\infty, \widehat{\theta}_n + \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}], \text{ or}$$
$$CI_n = [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}, \widehat{\theta}_n + \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}], \tag{2.2}$$

respectively.

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

Post-consistent model selection example. We consider a subsampling 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.

The model is

$$X_i = \theta + U_i, \quad \text{where } U_i \sim \text{i.i.d. } N(0, 1) \text{ for } i = 1, \ldots, n. \tag{2.3}$$

For the model selection problem, model 2 takes $\theta = 0$ and model 3 takes $\theta \in R$. Model selection is carried out using a likelihood ratio test that selects model 2 if $n^{1/2}|\overline{X}_n| \leq \kappa_n$ and model 3 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 2 is chosen with probability that goes to one as $n \to \infty$, and when $\theta_0 \neq 0$, model 3 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 asymptotic size of subsampling CIs based on post-conservative model selection procedures, such as AIC, is determined in Andrews and Guggenberger (forthcoming-a). It is far from the nominal level, but does not equal zero at least under some restrictions on a correlation matrix that arises.) The post-model selection estimator of θ_0 equals zero if model 2 is selected and $\overline{X}_n$ if model 3 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

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

$$T_n(\theta_0) = |n^{1/2}(\widehat{\theta}_n - \theta_0)|, \tag{2.4}$$

$\kappa_n > 0$, 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 defined in (2.1) with $T_n(\theta_0)$ defined in (2.4) and $c_{1-\alpha}$ defined below. In the case where $c_{1-\alpha}$ does not depend on θ_0, the CI is given by the third equation in (2.2).

2.2. Critical values and asymptotic size

We consider FCVs and subsampling critical values. The results below allow $c_{Fix}(1 - \alpha)$ to be any constant. Often, however, one takes

$$c_{Fix}(1 - \alpha) = c_\infty(1 - \alpha), \tag{2.5}$$

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 the true parameter is fixed and is not a point of discontinuity of the asymptotic

distribution of $T_n(\theta_0)$; see Section 3. In the post-consistent model selection example this corresponds to the limit distribution of the test statistic when the true θ_0 is fixed and different from zero. 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_d^2 distribution, denoted $\chi_d^2(1-\alpha)$.

To define subsampling critical values, let $\{b_n : n \geq 1\}$ be a sequence of subsample sizes. For brevity, we often write b_n as b. Let $\{\widehat{T}_{n,b,j} : j = 1, \ldots, q_n\}$ be subsample statistics defined below that are based primarily on subsamples of size b rather than the full sample. With i.i.d. observations, there are $q_n = n!/((n - b)!b!)$ different subsamples of size b and $\widehat{T}_{n,b,j}$ is determined primarily by the observations in the jth such subsample. With time series observations, say $\{X_1, \ldots, X_n\}$, there are $q_n = n-b+1$ subsamples of b consecutive observations, e.g., $Y_j = \{X_j, \ldots, X_{j+b-1}\}$, 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 $\{\widehat{T}_{n,b,j} : j = 1, \ldots, q_n\}$. They are defined by

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

$$c_{n,b}(1-\alpha) = \inf\{x \in R : L_{n,b}(x) \geq 1 - \alpha\}, \tag{2.6}$$

where $1(\cdot)$ denotes the indicator function, and they may depend on θ_0.

The subsample statistics $\{\widehat{T}_{n,b,j} : j = 1, \ldots, q_n\}$ are defined as follows. Let $\{T_{n,b,j}(\theta_0) : j = 1, \ldots, q_n\}$ be subsample statistics that are defined just as $T_n(\theta_0)$ is defined, but based on subsamples of size b 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,

(i) $T_{n,b,j}(\theta_0) = \tau_b(\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}$,   or

(ii) $T_{n,b,j}(\theta_0) = -\tau_b(\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}$,   or

(iii) $T_{n,b,j}(\theta_0) = |\tau_b(\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}|$. $\tag{2.7}$

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

$$U_{n,b}(x, \theta_0) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta_0) \leq x). \tag{2.8}$$

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

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

In some cases, the subsample statistics are defined to satisfy:

Assumption Sub2 $(\widehat{T}_{n,b,j} = T_{n,b,j}(\theta_0)$ for all $j \leq q_n)$.
    Note that $c_{n,b}(1 - \alpha)$ depends on the hypothesized parameter value θ_0 under Assumption Sub2, but not under Assumption Sub1. (Of course, the distribution of $c_{n,b}(1 - \alpha)$ may depend on the true parameter under Assumption Sub1 or Assumption Sub2.)

The distribution of the data is determined by a parameter γ of which θ is a sub-vector. Let Γ denote the parameter space for γ. The coverage probability of the CI defined in (2.1) when γ is the true parameter vector is

$$P_\gamma(\theta \in CI_n) = P_\gamma(T_n(\theta) \leq c_{1-\alpha}) = 1 - RP_n(\gamma), \tag{2.9}$$

where $RP_n(\gamma) = P_\gamma(T_n(\theta) > c_{1-\alpha})$. The exact (i.e., finite-sample) and asymptotic confidence sizes of CI_n are

$$ExCS_n = \inf_{\gamma \in \Gamma}(1 - RP_n(\gamma)) \quad \text{and} \quad AsyCS = \liminf_{n \to \infty} ExCS_n, \tag{2.10}$$

respectively.

Post-consistent model selection example (cont.). The subsampling critical values in this example are given by $c_{n,b}(1 - \alpha)$ obtained from the subsample statistics $\{T_{n,b,j}(\widehat{\theta}_n) : j = 1, \ldots, q_n\}$ defined in equation (iii) of (2.7) with $\widehat{\sigma}_{n,b,j} = 1$. Note that Assumption Sub1 holds. (The results given below also hold if Assumption Sub2 holds.)

3. Assumptions

3.1. Motivational example

In this section, we introduce the general assumptions under which our results hold. These assumptions allow for test statistics whose asymptotic distributions exhibit a type of discontinuity. The running example, which is a very simple post-consistent model selection example, is not sufficiently complex to illustrate the complexities that arise in many examples. In consequence, to illustrate the types of statistics that we want to cover, the type of discontinuity of interest, and the complexities that often arise, we start this section by describing a more complex example. After doing this, we introduce the general assumptions.

The example is a simple version of the example of inference in the linear instrumental variables model when instruments are potentially weak discussed in Andrews and Guggenberger (forthcoming-d) (AG hereafter). The model is given by a structural equation and a reduced-form equation

$$y_1 = y_2\theta + u, \qquad y_2 = z\pi + v, \tag{3.1}$$

where $y_1, y_2, z \in R^n$ and $\theta, \pi \in R$ are unknown parameters. Assume $\{(u_i, v_i, z_i) : i \leq n\}$ are i.i.d. with distribution F, where a subscript i denotes the ith component of a vector. The goal is to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. The test is based on the t statistic $T_n(\theta_0) = |n^{1/2}(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n|$, where $\widehat{\theta}_n = (y_2'P_zy_2)^{-1}y_2'P_zy_1$, $\widehat{\sigma}_n = \widehat{\sigma}_u(n^{-1}y_2'P_zy_2)^{-1/2}$, $\widehat{\sigma}_u^2 = (n - 1)^{-1}(y_1 - y_2\widehat{\theta}_n)'(y_1 - y_2\widehat{\theta}_n)$, and $P_z = zz'/z'z$. Define nuisance parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ by

$$\gamma_1 = |(E_Fz_i^2)^{1/2}\pi/\sigma_v|, \quad \gamma_2 = \rho, \quad \text{and} \quad \gamma_3 = (F, \pi), \quad \text{where}$$

$$\sigma_v^2 = E_Fv_i^2, \sigma_u^2 = E_Fu_i^2, \quad \text{and} \quad \rho = Corr_F(u_i, v_i). \tag{3.2}$$

The parameter spaces for γ_1 and γ_2 are $\Gamma_1 = \{x \in R : x \geq 0\}$ and $\Gamma_2 = [-1, 1]$. The details for the restrictions on the parameter space $\Gamma_3 = \Gamma_3(\gamma_1, \gamma_2)$ for γ_3 are given in AG and are such that the following central limit theorem (CLT) holds under sequences $\gamma = \gamma_n$ for which $\gamma_2 = \gamma_{2,n} \to h_2$:

$$\begin{pmatrix} (n^{-1}z'z)^{-1/2}n^{-1/2}z'u/\sigma_u \\ (n^{-1}z'z)^{-1/2}n^{-1/2}z'v/\sigma_v \end{pmatrix} \to_d \begin{pmatrix} \psi_{u,h_2} \\ \psi_{v,h_2} \end{pmatrix}$$

$$\sim N\left(0, \begin{pmatrix} 1 & h_2 \\ h_2 & 1 \end{pmatrix}\right). \tag{3.3}$$

In this example, the asymptotic distribution of the statistic $T_n(\theta_0)$ has a discontinuity at $\gamma_1 = 0$. Under different sequences $\gamma_1 = \gamma_{1,n}$ such that $\gamma_{1,n} \to 0$, the limit distribution of $T_n(\theta_0)$ may be different. More precisely, denote by $\gamma_{n,h}$ a sequence of nuisance

parameters $\gamma = \gamma_n$ such that $n^{1/2}\gamma_1 \to h_1$ and $\gamma_2 \to h_2$ and $h = (h_1, h_2)$. It is shown below that under $\gamma_{n,h}$, the limit distribution of $T_n(\theta_0)$ depends on h_1 and h_2 and only on h_1 and h_2. As long as h_1 is finite, the sequence γ_1 converges to zero, yet the limit distribution of $T_n(\theta_0)$ does not only depend on the limit point 0 of γ_1, but depends on how precisely γ_1 converges to zero, indexed by the convergence speed $n^{1/2}$ and the localization parameter h_1. In contrast, the limit distribution of $T_n(\theta_0)$ only depends on the limit point h_2 of γ_2 but not on how γ_2 converges to h_2. In that sense, the limit distribution is discontinuous in γ_1 at 0, but continuous on Γ_2 in γ_2. The parameter γ_3 does not influence the limit distribution of $T_n(\theta_0)$ by virtue of the CLT in (3.3).

If $h_1 < \infty$, it is shown in AG that under $\gamma_{n,h}$

$$
\begin{pmatrix} y_2'P_z u/(\sigma_u\sigma_v) \\ y_2'P_z y_2/\sigma_v^2 \\ \widehat{\sigma}_u^2/\sigma_u^2 \end{pmatrix} \to_d \begin{pmatrix} \xi_{1,h} \\ \xi_{2,h} \\ \eta_{u,h}^2 \end{pmatrix}
$$

$$
= \begin{pmatrix} (\psi_{v,h_2} + h_1)\psi_{u,h_2} \\ (\psi_{v,h_2} + h_1)^2 \\ (1 - h_2\xi_{1,h}/\xi_{2,h})^2 + (1 - h_2^2)\xi_{1,h}^2/\xi_{2,h}^2 \end{pmatrix} \tag{3.4}
$$

and thus $T_n(\theta_0) \to_d |\xi_{1,h}/(\xi_{2,h}\eta_{u,h}^2)^{1/2}|$. If $h_1 = \infty$, $T_n(\theta_0) \to_d |\eta_\infty|$, where η_∞ has a standard normal limit distribution that does not depend on h_2.

3.2. Parameter space

We now return to the general case. The parameter γ has up to three components: $\gamma = (\gamma_1, \gamma_2, \gamma_3) = ((\theta_1', \eta_1')', (\theta_2', \eta_2')', \gamma_3)$, where $\theta = (\theta_1', \theta_2')'$, $\eta = (\eta_1', \eta_2')'$, $\theta_j \in R^{d_j}$ for $j = 1, 2$, and $\eta_j \in R^{s_j}$ for $j = 1, 2$. Points of discontinuity of the asymptotic distribution of the test statistic of interest are determined by the first component, $\gamma_1 \in R^p$, which may contain part of the parameter of interest, viz., θ_1. Through reparameterization 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, $\gamma_2 (\in R^q)$, 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 component γ_2 may contain part of the parameter of interest, θ_2. 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 $s_2 = 0$). The parameter space for γ_2 is $\Gamma_2 \subset R^q$.

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 $\mathcal{T}_3$ and hence can be finite or infinite dimensional. For example, error distributions can be included in γ_3. The parameter space for γ_3 is $\Gamma_3(\gamma_1, \gamma_2)(\subset \mathcal{T}_3)$, which may depend on γ_1 and γ_2.

Assumption A1. The parameter space for γ is

$$
\Gamma = \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\}. \tag{3.5}
$$

Post-consistent model selection example (cont.). In this example, no parameters $\gamma_2, \gamma_3, \theta_2$, or η appear. Assumption A1 holds with $\gamma = \gamma_1 = \theta = \theta_1 \in R$, $p = d = d_1 = 1$, $d_2 = 0$, and $\Gamma = \Gamma_1 = \Theta = R$.

3.3. Convergence assumption

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 of a distribution G by $q(1 - \alpha) = \inf\{x \in R : G(x) \geq 1 - \alpha\}$. Let $G(x-) = \lim_{\varepsilon \searrow 0} G(x - \varepsilon)$, where "$\lim_{\varepsilon \searrow 0}$" denotes the limit as $\varepsilon > 0$ declines to zero. The distributions J_h and J_{h^0} considered below are distributions of proper random variables that are finite with probability one. All limits are as $n \to \infty$. For a sequence of constants $\{\kappa_n : n \geq 1\}$, let $\kappa_n \to [\kappa_{1,\infty}, \kappa_{2,\infty}]$ denote that $\kappa_{1,\infty} \leq \liminf_{n\to\infty} \kappa_n \leq \limsup_{n\to\infty} \kappa_n \leq \kappa_{2,\infty}$.

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 (forthcoming-a) $r = 1$. In a given example, the value of r is determined such that under sequences γ_1 satisfying $n^r\gamma_1 \to h_1$ we obtain sequences of distributions that are contiguous to the distribution at a point of discontinuity of the asymptotic distribution.

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

$$
H = \{h = (h_1, h_2) \in R_\infty^{p+q} : \exists \{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3})
$$

$$
\in \Gamma : n \geq 1\}\text{such that } n^r\gamma_{n,1} \to h_1 \text{ and } \gamma_{n,2} \to h_2\}, \tag{3.6}
$$

where $R_\infty = R \cup \{\pm\infty\}$ and $R_\infty^{p+q} = R_\infty \times \cdots \times R_\infty$ (with $p + q$ copies). 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.

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 Assumption B1 below. For a given model, there is a single fixed $r > 0$. In addition, the limit distribution under $\{\gamma_{n,h} : n \geq 1\}$ of the test statistic of interest does not depend on $\gamma_{n,h,3}$, so we do not make the dependence of $\gamma_{n,h}$ on $\gamma_{n,h,3}$ explicit.

Given any $h = (h_1, h_2) \in H$, define $h^0 = (0, h_2)$. Let 0_p denote a p vector of zeros.

We use the following assumption.

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, $T_n(\theta_{n,h}) \to_d J_h$ under $\{\gamma_{n,h} : n \geq 1\}$, 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 $\theta_{n,h} = (\theta_{n,h,1}', \theta_{n,h,2}')'$, and (ii) for all sequences $\{\gamma_{n,h^0} : n \geq 1\}$ and some distribution J_{h^0}, $T_n(\theta_{n,h^0}) \to_d J_{h^0}$ under $\{\gamma_{n,h^0} : n \geq 1\}$, where $\gamma_{n,h^0} = (\gamma_{n,h^0,1}, \gamma_{n,h^0,2}, \gamma_{n,h^0,3}) = (0_p, (\theta_{n,h^0,2}', \eta_{n,h^0,2}')', \gamma_{n,h^0,3})$ and $\theta_{n,h^0} = (0_p', \theta_{n,h^0,2}')'$.

Assumption B1 models discontinuity in the asymptotic distribution of $T_n(\theta_{n,h})$ at $\gamma_1 = 0$ when J_h and J_{h^0} differ because the sequences $\{\gamma_{n,h} : n \geq 1\}$ and $\{\gamma_{n,h^0} : n \geq 1\}$ both converge to $(0_p, h_2, \gamma_3)$ when $\gamma_{n,h,3} \to \gamma_3$ yet the asymptotic distributions J_h and J_{h^0} differ. If $\gamma_{n,h}$ does not depend on n (which necessarily requires $h_1 = 0$ because $\|h\| < \infty$), Assumption B1(i) is a standard assumption in the subsampling 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 B1 is a weaker assumption than Assumption B, which is employed in Andrews and Guggenberger (forthcoming-c) (provided there exists an $h \in H \cap R^{p+q}$ such that $h^0 \in H$, which

holds quite generally). For example, Assumption B1 holds in the consistent model selection example considered here, but Assumption B does not because the latter requires convergence of $T_n(\theta_{n,h})$ to the same distribution J_∞ for all sequences $\{\gamma_{n,h} : n \geq 1\}$ for which $h = \infty$. The latter fails; see below.

Post-consistent model selection example (cont.). In this example, we take $r = 1/2$ and $\gamma_{n,h} (=\theta_{n,h}) = hn^{-1/2}$, where $h \in R$, in Assumption B1. We now 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

$$P_{\gamma_n}(n^{1/2}|\overline{X}_n| \leq \kappa_n) = P_{\gamma_n}(|n^{1/2}(\overline{X}_n - \theta_n) + n^{1/2}\theta_n| \leq \kappa_n)$$
$$= P_{\gamma_n}(|O_p(1) + O(1)| \leq \kappa_n) \to 1 \quad \text{and}$$

$$P_{\gamma_n}(\widehat{\theta}_n = a\overline{X}_n) \to 1, \tag{3.7}$$

where the second equality uses the fact that $n^{1/2}(\overline{X}_n - \theta_n) \sim N(0, 1)$ and the second convergence result uses the definition of $\widehat{\theta}_n$ in (2.4).

Hence when the true value is $\theta_{n,h}$, $\widehat{\theta}_n = a\overline{X}_n$ wp $\to 1$, and we have wp $\to 1$ under $\theta_{n,h}$,

$$
\begin{aligned}
T_n(\theta_{n,h}) &= |n^{1/2}(a\overline{X}_n - \theta_{n,h})| \\
&= |an^{1/2}(\overline{X}_n - \theta_{n,h}) + (a-1)h| \\
&\sim |aZ + (a-1)h| \sim J_h, \quad \text{where } Z \sim N(0, 1) \quad \text{and} \quad (3.8)
\end{aligned}
$$

$$J_h(x) = \begin{cases} \Phi(a^{-1}(x + (1-a)h)) \\ \quad -\Phi(a^{-1}(-x + (1-a)h)) & \text{if } a \in (0, 1) \\ 1(x \geq |h|) & \text{if } a = 0, \end{cases}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. 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 $c_{h^0}(1-\alpha) = c_0(1-\alpha) = az_{1-\alpha/2}$.

Eq. (3.8) implies that Assumption B1(i) holds. For any sequence $\{\gamma_{n,h^0} : n \geq 1\}$ as in Assumption B1(ii), we have $n^{1/2}\gamma_{n,h^0} = O(1)$, (3.7) holds, and (3.8) holds with γ_{n,h^0} in place of $\gamma_{n,h} (=\theta_{n,h})$. Hence, Assumption B1(ii) holds with $J_{h^0}(x) = J_0(x)$.

Assumption B of Andrews and Guggenberger (forthcoming-c) fails in this example because, as is obvious and known, the asymptotic distribution of $T_n(\theta_{n,h})$ (when it exists) differs for a sequence $\{\theta_{n,h} : n \geq 1\}$ that converges to 0 but slowly enough that $n^{1/2}|\overline{X}_n| > \kappa_n$ occurs with probability that is bounded away from 0 and 1 from a sequence $\{\theta_{n,h} : n \geq 1\}$ for which $n^{1/2}|\overline{X}_n| > \kappa_n$ occurs wp $\to 1$. For both such sequences, $h = \infty$.

3.4. Subsampling assumptions

To determine the asymptotic coverage probabilities of FCV CIs, the assumptions above are all that are needed. For subsampling CIs, we require the following additional assumptions[7]:

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

Assumption D. (i) $\{T_{n,b,j}(\theta) : j = 1, \dots, q_n\}$ are identically distributed under any $\gamma \in \Gamma$ for all $n \geq 1$, where $\gamma = (\gamma_1, \gamma_2, \gamma_3) = ((\theta_1', \eta_1')', (\theta_2', \eta_2')', \gamma_3)$ and $\theta = (\theta_1', \theta_2')'$, and (ii) $T_{n,b,j}(\theta)$ and $T_b(\theta)$ have the same distribution under any $\gamma \in \Gamma$ for all $n \geq 1$.

---

[7] Assumptions that are not indexed by "1" are the same as assumptions in Andrews and Guggenberger (forthcoming-c). Assumptions that are indexed by "1" concern the same quantities as, but are different from, corresponding assumptions in Andrews and Guggenberger (forthcoming-c) that are not indexed by "1."

Assumption E1. For the sequence $\{\gamma_{n,h} : n \geq 1\}$ in Assumption B1(i), $U_{n,b}(x, \theta_{n,h}) - E_{\gamma_n}U_{n,b}(x, \theta_{n,h}) \to_p 0$ under $\{\gamma_{n,h} : n \geq 1\}$ 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 G1. For the sequence $\{\gamma_{n,h} : n \geq 1\}$ in Assumption B1(i), $L_{n,b}(x) - U_{n,b}(x, \theta_{n,h}) \to_p 0$ for all $x \in C(J_{h^0})$ under $\{\gamma_{n,h} : n \geq 1\}$.

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

Assumption E1 holds automatically for subsample statistics that are defined as above when the observations are i.i.d. for each fixed $\gamma \in \Gamma$ (by a U-statistic inequality of Hoeffding using the same argument as in PRW, p. 44). For stationary strong mixing observations, Assumption E1 holds provided

$$\sup_{\gamma \in \Gamma} \alpha_\gamma(m) \to 0 \quad \text{as } m \to \infty, \tag{3.9}$$

where $\{\alpha_\gamma(m) : m \geq 1\}$ are the strong mixing numbers of the observations when the true parameter is γ. This follows by the same argument as given in PRW, pp. 71–72 (which establishes L^2 convergence using a strong mixing covariance bound).

Assumption F1 is designed to avoid the requirement that $J_h(x)$ is continuous in x because this assumption is violated in some examples, such as the consistent model selection example, for some values of h and some values of x. Assumption F1 holds if either (i) $J_{h^0}(x)$ is continuous and 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 in the post-consistent model selection estimator with constant $a = 0$, then condition (i) fails, but condition (ii) holds.

Assumption G1 holds automatically when $\{\widehat{T}_{n,b,j}\}$ satisfy Assumption Sub2. To verify that Assumption G1 holds when $\{\widehat{T}_{n,b,j}\}$ 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. $\tau_b/\tau_n \to 0$.

This is a standard assumption in the subsampling 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 $\tau_b/\tau_n = (b/n)^s \to 0$.

Lemma 1. *Assumptions* t1, t2, Sub1, A1, B1, C, D *and* H *imply Assumption* G1.

Comment. Lemma 1 is a special case of Lemma 5, which does not impose Assumption t2 and, hence, covers studentized and non-studentized t statistics. Lemma 5 is stated in the Appendix for expositional convenience.

Post-consistent model selection example (cont.). We now verify Assumptions C, D, E1–G1, and H for this example for an arbitrary choice of the parameter h. We choose $\{b = b_n : n \geq 1\}$ so that Assumption C holds, Assumption D and E1 hold because the observations are i.i.d. for each fixed $\theta \in R$, Assumption H holds because $\tau_b/\tau_n = b^{1/2}/n^{1/2} \to 0$ by Assumption C, and Assumption G1 holds by Lemma 1 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 $c_{h^0}(1-\alpha) = az_{1-\alpha/2}$.

4. Asymptotic results

The main result of this paper concerns the asymptotic behavior of FCV and subsampling CIs under a sequence $\{\gamma_{n,h} : n \geq 1\}$.

Theorem 2. (a) *Suppose Assumption B1 (i) holds. Then, $P_{\gamma_{n,h}}(T_n(\theta_{n,h})$
$\leq c_{Fix}(1 - \alpha)) \rightarrow [J_h(c_{Fix}(1 - \alpha)-), J_h(c_{Fix}(1 - \alpha))]$.*

(b) *Suppose Assumptions A1, B1, C, D, and E1–G1 hold. Then,*
$P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_{n,b}(1 - \alpha)) \rightarrow [J_h(c_{h^0}(1 - \alpha)-), J_h(c_{h^0}(1 - \alpha))]$.

Comments. 1. If $J_h(c_{h^0}(1 - \alpha)) < 1 - \alpha$, then part (b) shows that the subsampling CI has asymptotic confidence size less than its nominal level $1 - \alpha$: $AsyCS < 1 - \alpha$. If $J_h(c_{h^0}(1 - \alpha)-) > 1 - \alpha$, then the subsampling CI is not asymptotically similar. Analogous statements apply to FCV CIs with $J_h(c_{Fix}(1 - \alpha))$ in place of $J_h(c_{h^0}(1 - \alpha))$.

2. If $J_h(x)$ is continuous at $x = c_{h^0}(1 - \alpha)$, then the result of Theorem 2(b) becomes $P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_{n,b}(1 - \alpha)) \rightarrow J_h(c_{h^0}(1 - \alpha))$.

3. Typically Assumption B1(i) holds for an infinite number of values h, say $h \in H^*(\subset R^p)$. In this case, Comments 1 and 2 apply for any $h \in H^*$.

Post-consistent model selection example (cont.). For $a = 0$, Theorem 2(b) implies that the limit of the coverage probability of the subsampling CI under $\gamma_{n,h} (=\theta_{n,h}) = hn^{-1/2}$ is

$$J_h(c_{h^0}(1 - \alpha)) = J_h(0) = 1(0 \geq |h|) = 0 \quad \text{for } |h| > 0. \tag{4.1}$$

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

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

$$J_h(c_{h^0}(1 - \alpha)) = J_h(az_{1-\alpha/2}). \tag{4.2}$$

Using (3.8), for $a \in (0, 1)$, we have

$$\lim_{h \to \infty} J_h(az_{1-\alpha/2}) = 0. \tag{4.3}$$

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

Fig. 1 graphs the asymptotic coverage probability of the nominal 95% subsampling CI under $\gamma_{n,h}$ as a function of $|h|$ for various values of a, namely $a = 0, .25, .5,$ and $.75$. The results are obtained by simulation from (4.2) using 100,000 simulation repetitions. Fig. 1 illustrates how the degree of under-coverage of the subsampling CI increases as a decreases and as $|h|$ increases. In the extreme case of $a = 0$, the asymptotic coverage probability equals zero for all $|h| > 0$. For any positive value a considered, the asymptotic coverage probability equals the nominal level $.95$ when $|h| = 0$, decreases as $|h|$ increases, and approaches zero as $|h| \to \infty$.

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 (forthcoming-a,d) do not work in this example because Assumptions LF, LS, and LH in the Appendix of Andrews and Guggenberger (forthcoming-d) fail. (For example, Assumption LF fails when $a = 0$ because $H = R_\infty$, $c_h(1 - \alpha) = |h|$, and $\sup_{h \in R_\infty} |h| = \infty$.) Andrews and Guggenberger (forthcoming-a) does provide size-correction methods for CIs based on post-*conservative* model selection estimators.
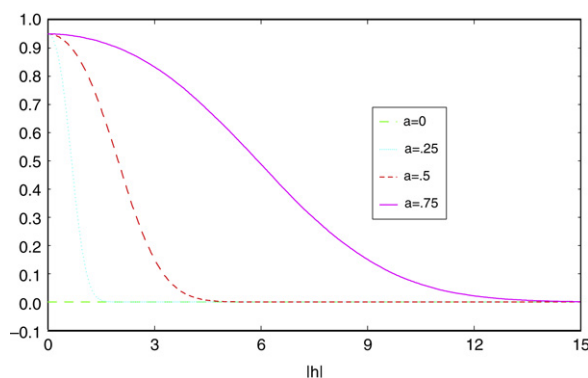

Fig. 1. Asymptotic coverage probabilities as a function of $|h|$ for several values of the constant a.

5. Proofs

The following Lemmas are used in the proof of Theorem 2.

Lemma 3. *Suppose* (i) *for some df's $L_n(\cdot)$ and $G_L(\cdot)$ on R, $L_n(x) \rightarrow_p G_L(x)$ for all $x \in C(G_L)$,* (ii) *$T_n \rightarrow_d G_T$, 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 $c_n := \inf\{x \in R : L_n(x) \geq 1 - \alpha\}$,* (a) *$c_n \rightarrow_p c_\infty$ and* (b) *$P(T_n \leq c_n) \rightarrow [G_T(c_\infty-), G_T(c_\infty)]$.*

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

2. Lemma 3 is the same as Lemma 5 of Andrews and Guggenberger (forthcoming-c). For completeness, we repeat its proof below.

Lemma 4. *Suppose Assumptions A1, B1, C, D, and E1–G1 hold. Let $\{\gamma_{n,h} : n \geq 1\}$ be as in Assumption B1 (i). Then, under $\{\gamma_{n,h} : n \geq 1\}$, we have*

(a) *$E_{\gamma_{n,h}} U_{n,b}(x, \theta_{n,h}) \rightarrow J_{h^0}(x)$ for all $x \in C(J_{h^0})$,*
(b) *$U_{n,b}(x, \theta_{n,h}) \rightarrow_p J_{h^0}(x)$ for all $x \in C(J_{h^0})$,*
(c) *$L_{n,b}(x) \rightarrow_p J_{h^0}(x)$ for all $x \in C(J_{h^0})$,*
(d) *$c_{n,b}(1 - \alpha) \rightarrow_p c_{h^0}(1 - \alpha)$, and*
(e) *$P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_{n,b}(1 - \alpha)) \rightarrow [J_h(c_{h^0}(1 - \alpha)-), J_h(c_{h^0}(1 - \alpha))]$.*

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

$$L_n(c_\infty - \varepsilon) \rightarrow_p G_L(c_\infty - \varepsilon) < 1 - \alpha \quad \text{and}$$
$$L_n(c_\infty + \varepsilon) \rightarrow_p G_L(c_\infty + \varepsilon) > 1 - \alpha \tag{5.4}$$

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

$$P(A_n(\varepsilon)) \rightarrow 1, \quad \text{where } A_n(\varepsilon) = \{c_\infty - \varepsilon \leq c_n \leq c_\infty + \varepsilon\}. \tag{5.5}$$

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

$$P(T_n \leq c_\infty - \varepsilon, A_n(\varepsilon)) \leq P(T_n \leq c_n, A_n(\varepsilon)) \leq P(T_n \leq c_\infty + \varepsilon). \tag{5.6}$$

Hence,

$$\limsup_{n\to\infty} P(T_n \leq c_n) = \limsup_{n\to\infty} P(T_n \leq c_n, A_n(\varepsilon))$$
$$\leq \limsup_{n\to\infty} P(T_n \leq c_\infty + \varepsilon) = G_T(c_\infty + \varepsilon), \quad \text{and}$$
$$\liminf_{n\to\infty} P(T_n \leq c_n) = \liminf_{n\to\infty} P(T_n \leq c_n, A_n(\varepsilon))$$
$$\geq \liminf_{n\to\infty} P(T_n \leq c_\infty - \varepsilon, A_n(\varepsilon)) = G_T(c_\infty - \varepsilon) \tag{5.7}$$

using assumption (ii), $c_\infty \pm \varepsilon \in C(G_T)$, and (5.5). Given a sequence $\{\varepsilon_k : k \geq 1\}$ as above, (5.7) establishes part (b). □

Proof of Lemma 4. First, we prove part (a). The proof is similar to that of Lemma 6(ii) of Andrews and Guggenberger (forthcoming-c). We have

$$E_{\gamma_{n,h}} U_{n,b}(x, \theta_{n,h}) = q_n^{-1} \sum_{j=1}^{q_n} P_{\gamma_{n,h}}(T_{n,b,j}(\theta_{n,h}) \leq x)$$
$$= P_{\gamma_{n,h}}(T_{n,b,1}(\theta_{n,h}) \leq x) = P_{\gamma_{n,h}}(T_b(\theta_{n,h}) \leq x), \tag{5.8}$$

where the first equality holds by definition of $U_{n,b}(x, \theta_{n,h})$, the second equality holds by Assumption D(i), and the last equality holds by Assumption D(ii).

We now show that $P_{\gamma_{n,h}}(T_{b_n}(\theta_{n,h}) \leq x) \to J_{h^0}(x)$ for all $x \in C(J_{h^0})$ by showing that any subsequence $\{t_n\}$ of $\{n\}$ has a sub-subsequence $\{s_n\}$ for which $P_{\gamma_{s_n,h}}(T_{b_{s_n}}(\theta_{s_n,h}) \leq x) \to J_{h^0}(x)$ (where θ_{s_n} is a sub-vector of γ_{s_n}). 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 C(i). Because $\{b_{s_n}\}$ is strictly increasing, it is a subsequence of $\{n\}$.

Below we show that Assumption B1(ii) implies that for any subsequence $\{u_n\}$ of $\{n\}$ and any sequence $\{\gamma_{u_n}^* = (\gamma_{u_n,1}^*, \gamma_{u_n,2}^*, \gamma_{u_n,3}^*) \in \Gamma : n \geq 1\}$ that satisfies (i) $u_n^r \gamma_{u_n,1}^* \to 0$ and (ii) $\gamma_{u_n,2}^* \to h_2 \in R^q$, we have

$$P_{\gamma_{u_n}^*}(T_{u_n}(\theta_{u_n}^*) \leq y) \to J_{h^0}(y), \tag{5.9}$$

for all $y \in C(J_{h^0})$ (where $\theta_{u_n}^*$ is a sub-vector of $\gamma_{u_n}^*$). We apply this result with $u_n = b_{s_n}$, $\gamma_{u_n}^* = \gamma_{s_n,h}$, and $y = x$ to obtain the desired result $P_{\gamma_{s_n,h}}(T_{b_{s_n}}(\theta_{s_n,h}) \leq x) \to J_{h^0}(x)$, where (i) and (ii) hold by the properties of $\{\gamma_{n,h} : n \geq 1\}$.

For the proof of part (a), it remains to show (5.9). Because $h^0 \in H$ (by Assumption B1(i)), 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 $k^r \gamma_{k,1}^+ \to 0$ and $\gamma_{k,2}^+ \to h_2$ 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. For $n \geq 1$, 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 $k^r \gamma_{k,1}^{**} \to 0$ and $\gamma_{k,2}^{**} \to h_2$ as $k \to \infty$. Hence, $\{\gamma_k^{**} : k \geq 1\}$ is of the form $\{\gamma_{n,h^0} : n \geq 1\}$ and Assumption B1(ii) implies that $P_{\gamma_k^{**}}(T_k(\theta_k^{**}) \leq y) \to J_{h^0}(y)$ for all $y \in C(J_{h^0})$ (where θ_k^{**} is a sub-vector of γ_k^{**}). Because $\{u_n\}$ is a subsequence of $\{k\}$ and $\gamma_k^{**} = \gamma_{u_n}^*$ when $k = u_n$, the latter implies that $P_{\gamma_{u_n}^*}(T_{u_n}(\theta_{u_n}^*) \leq y) \to J_{h^0}(y)$, as desired.

Part (b) follows from part (a) and Assumption E1. Part (c) follows from part (b) and Assumption G1. Parts (d) and (e) are established by applying Lemma 3 with $L_n(x) = L_{n,b}(x)$ and $T_n = T_n(\theta_{n,h})$ and verifying the conditions of Lemma 3 using (i) part (c), (ii) $T_n(\theta_{n,h}) \to_d J_h$ under $\{\gamma_{n,h} : n \geq 1\}$ (by Assumption B1(i)), and (iii) Assumption F1. □

Proof of Theorem 2. 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(e). □

We now provide sufficient conditions for Assumption G1 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 $T_{n,b,j}(\theta_0)$ is defined, let $(\widehat{\theta}_{n,b,j}, \widehat{\sigma}_{n,b,j})$ be the subsample statistics that are defined exactly as $(\widehat{\theta}_n, \widehat{\sigma}_n)$ are defined, but based on the jth subsample of size b. In analogy to $U_{n,b}(x, \theta_{n,h})$ defined in (2.8), we define

$$U_{n,b}^\sigma(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(d_b \widehat{\sigma}_{n,b,j} \leq x) \tag{5.10}$$

for a sequence of normalization constants $\{d_n : n \geq 1\}$ (for which Assumption BB1 below holds). Although $U_{n,b}^\sigma(x)$ depends on $\{d_n : n \geq 1\}$, we suppress the dependence for notational simplicity.

We now state modified versions of Assumptions B1, D, E1 and H that are used with studentized statistics when Assumption Sub1 holds.

Assumption BB1. For r, h, h^0, and $\{\gamma_{n,h} : n \geq 1\}$ as in Assumption B1(i) and for some distribution (V_h, W_h) on R^2, $(a_n(\widehat{\theta}_n - \theta_{n,h}), d_n \widehat{\sigma}_n) \to_d (V_h, W_h)$ under $\{\gamma_{n,h} : n \geq 1\}$, 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 $\theta_{n,h} = (\theta_{n,h,1}', \theta_{n,h,2}')'$, (ii) $P_{\gamma_{n,h}}(\widehat{\sigma}_{n,b,j} > 0$ for all $j = 1, \ldots, q_n) \to 1$ under $\{\gamma_{n,h} : n \geq 1\}$, and (iii) $W_h(0) = 0$.

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

Assumption EE1. For the sequence $\{\gamma_{n,h} : n \geq 1\}$ in Assumption BB1(i) and the constants $\{d_n : n \geq 1\}$ in Assumption BB1(i), $U_{n,b}^\sigma(x) - E_{\gamma_{n,h}} U_{n,b}^\sigma(x) \to_p 0$ under $\{\gamma_{n,h} : n \geq 1\}$ for all $x \in R$.

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

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 $a_n = n^{1/2}$.

Assumption BB1 implies that $T_n(\theta_{n,h}) \to_d J_h$ in Assumption B1(i) with $\tau_n = a_n/d_n$ (by the continuous mapping theorem using Assumption BB1(iii)). Assumption DD implies Assumption D. Assumption DD is not restrictive given the standard methods of defining subsample statistics. Assumption EE1 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 (3.9) for each fixed $\gamma \in \Gamma$ provided the subsamples are constructed as described in Section 2.2 (for the same reason that Assumption E1 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 BB1(i)). For example, this occurs in a model with weak instruments, see Andrews and Guggenberger (forthcoming-d).

The following Lemma generalizes Lemma 1 because (i) the following Lemma does not impose Assumption t2 and (ii) Assumptions t1, t2, B1, D, and H imply Assumptions BB1, DD, EE1, and HH with $\widehat{\sigma}_{n,b,j} = d_n = 1$.

Lemma 5. *Assumptions* t1, Sub1, A1, B1, BB1, C, D, DD, E1, EE1, *and* HH *imply Assumption* G1.

Comment. The proof of Lemma 5 is a variant of those of Theorem 11.3.1(i) and 12.2.2(i) of PRW and Lemma 4 of Andrews and Guggenberger (forthcoming-c).

Proof of Lemma 5. We have $U_{n,b}(x, \theta_{n,h}) \to_p J_{h^0}(x)$ for all $x \in C(J_{h^0})$ under $\{\gamma_{n,h} : n \geq 1\}$ by Lemma 4(b) (which does not require Assumptions F1 and G1 in its proof). Define $R_n(t) := q_n^{-1} \sum_{j=1}^{q_n} 1(|\tau_b(\widehat{\theta}_n - \theta_{n,h})/\widehat{\sigma}_{n,b,j}| \geq t)$. Using

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

(5.11)

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

$$|\tau_b(\widehat{\theta}_n - \theta_{n,h})/\widehat{\sigma}_{n,b,j}| \geq t \text{ iff } (a_b/a_n)a_n|\widehat{\theta}_n - \theta_{n,h}| \geq d_b\widehat{\sigma}_{n,b,j}t \quad (5.12)$$

provided $\widehat{\sigma}_{n,b,j} > 0$, which holds uniformly in $j = 1, \ldots, q_n$ wp $\to 1$ by Assumption BB1(ii). By Assumption BB1(i) and Assumption HH, $(a_b/a_n)a_n|\widehat{\theta}_n - \theta_{n,h}| = o_p(1)$ under $\{\gamma_{n,h}\}$. Therefore, for any $\delta > 0$, $R_n(t) \leq q_n^{-1} \sum_{j=1}^{q_n} 1(\delta \geq d_b\widehat{\sigma}_{n,b,j}t) = U_{n,b}^\sigma(\delta/t)$, where the inequality holds wp $\to 1$. Now, by an argument as in the proof of Lemma 4(a) and (b) (which uses Assumption EE1, but does not use Assumption G1) applied to the statistic $d_n\widehat{\sigma}_n$ rather than $T_n(\theta_{n,h})$, we have $U_{n,b}^\sigma(x) \to_p W_{h^0}(x)$ for all $x \in C(W_{h^0})$ under $\{\gamma_{n,h}\}$. Therefore, $U_{n,b}^\sigma(\delta/t) \to_p W_{h^0}(\delta/t)$ for $\delta/t \in C(W_{h^0})$ under $\{\gamma_{n,h}\}$. By Assumption BB1(iii), W_{h^0} does not have positive mass at zero and, hence, $W_{h^0}(\delta/t) \to 0$ as $\delta \to 0$. We can therefore establish that $R_n(t) \to_p 0$ for any $t > 0$ by letting δ go to zero such that $\delta/t \in C(W_{h^0})$. $\square$

Acknowledgments

References

Andrews, D.W.K., Guggenberger, P., 2005. The limit of finite-sample size and a problem with subsampling. Cowles Foundation Discussion Paper No. 1605, Yale University.

Andrews, D.W.K., Guggenberger, P., 2009. Hybrid and size-corrected subsampling methods. Econometrica, vol. 77 (forthcoming-a).

Andrews, D.W.K., Guggenberger, P., 2009. Validity of subsampling and "plug-in asymptotic"inference for parameters defined by moment inequalities. Econometric Theory, vol. 25 (forthcoming-b).

Andrews, D.W.K., Guggenberger, P., 2010. Asymptotic size and a problem with subsampling and with the *m* out of *n* bootstrap. Econometric Theory, vol. 26 (forthcoming-c).

Andrews, D.W.K., Guggenberger, P., 2010. Applications of subsampling, hybrid, and size-correction methods. Journal of Econometrics (forthcoming-d).

Beran, R., 1982. Estimating sampling distributions: The bootstrap and competitors. Annals of Statistics 10, 212–225.

Guggenberger, P., 2008. The impact of a Hausman pretest on the size of hypothesis tests. Cowles Foundation Discussion Paper No. 1651, Yale University.

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

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

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

Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses, 3rd ed. Wiley, New York.

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

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

Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.

Pötscher, B.M., 2007. Confidence sets based on sparse estimators are necessarily large. Unpublished manuscript, Department of Statistics, University of Vienna.

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

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

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

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

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

Sherman, M., Carlstein, E., 1996. Replicate histograms. Journal of the American Statistical Association 91, 566–576.

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