Hybrid and Size-Corrected Subsampling Methods

Donald W. K. Andrews*
Cowles Foundation for Research in Economics
Yale University

Patrik Guggenberger
Department of Economics
UCLA

June 2005
Revised: October 2008

* Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-0417911 and SES-0751517. Guggenberger gratefully acknowledges research support from a faculty research grant from UCLA in 2005 and from the National Science Foundation via grant number SES-0748922. The authors thank three referees, the co-editor Whitney Newey, Hannes Leeb, Azeem Shaikh, Jim Stock, and the participants at a number of seminars and conferences at which the paper was presented.

Abstract

This paper considers inference in a broad class of non-regular models. The models considered are non-regular in the sense that standard test statistics have asymptotic distributions that are discontinuous in some parameters. It is shown in Andrews and Guggenberger (2009a) that standard fixed critical value, subsampling, and m out of n bootstrap methods often have incorrect asymptotic size in such models. This paper introduces general methods of constructing tests and confidence intervals that have correct asymptotic size. In particular, we consider a hybrid subsampling/fixed-critical-value method and size-correction methods. The paper discusses two examples in detail. They are: (i) confidence intervals in an autoregressive model with a root that may be close to unity and conditional heteroskedasticity of unknown form and (ii) tests and confidence intervals based on a post-conservative model selection estimator.

Keywords: Asymptotic size, autoregressive model, m out of n bootstrap, exact size, hybrid test, model selection, over-rejection, size correction, subsample, confidence interval, subsampling test.

JEL Classification Numbers: C12, C15.

1 Introduction

Non-regular models are becoming increasingly important in econometrics and statistics as developments in computation make it feasible to employ more complex models. In a variety of non-regular models, however, methods based on a standard asymptotic fixed critical value (FCV) or the bootstrap do not yield tests or confidence intervals with the correct size even asymptotically. In such cases, the usual prescription in the literature is to use subsampling or m out n bootstrap methods (where n denotes the sample size and m denotes the bootstrap sample size). For references, see Andrews and Guggenberger (2009a), hereafter denoted AG1. However, AG1 shows that in a fairly broad array of non-regular models these methods do not deliver correct asymptotic size (defined to be the limit of exact size). The purpose of this paper is to provide general methods of constructing tests and confidence intervals (CIs) that do have correct asymptotic size in such models.

The results cover cases in which a test statistic has an asymptotic distribution that is discontinuous in some parameters. Examples include inference for (i) post-conservative model-selection procedures (such as those based on AIC), (ii) parameters in scalar and vector autoregressive models with roots that may be close to unity, (iii) models with a parameter near a boundary, (iv) models with lack of identification at some point(s) in the parameter space, such as models with weak instruments, (v) predictive regression models with nearly-integrated regressors, (vi) threshold autoregressive models, (vii) tests of stochastic dominance, (viii) non-differentiable functions of parameters, and (ix) differentiable functions of parameters that have zero first-order derivative.

The methods considered here are quite general. However, their usefulness is greatest in models in which other methods, such as those based on a standard asymptotic FCV or the bootstrap, are not applicable. In models in which other methods work properly (in the sense that the limit of their exact size equals their nominal level), such methods are often preferable to the methods considered here in terms of the accuracy of the asymptotic approximations and/or the power of the test or length of the CI they generate.

The first method considered in the paper is a hybrid method that takes the critical value for a given test statistic to be the maximum of a subsampling critical value and the FCV that applies when the true parameters are not near a point of discontinuity of the asymptotic distribution. The latter is usually a normal or chi-square critical value. By simply taking the maximum of these two critical values, one obtains a test or CI that has correct asymptotic size in many cases where the FCV, subsampling, or both methods have incorrect asymptotic size. Furthermore, the paper shows that the hybrid method has the feature that relative to a subsampling method either (i) the subsampling method has correct size asymptotically and the subsampling and hybrid critical values are the same asymptotically or (ii) the subsampling method has incorrect size asymptotically and the hybrid method reduces the magnitude of over-rejection for at least some parameter values, sometimes eliminating size distortion.

The second method considered in the paper is a size-correction (SC) method. This method can be applied to FCV, subsampling, and hybrid procedures. The basic idea is to use the formulae given in AG1 for the asymptotic sizes of these procedures and to increase the magnitudes of the critical values (by adding a constant or reducing the nominal level) to achieve a test whose asymptotic size equals the desired asymptotic level. Closed form solutions are obtained for the SC values (based on adding a constant). Numerical work in a number of different examples shows that computation of the SC values is tractable.

The paper provides analytical comparisons of the asymptotic power of different SC tests and finds that the SC hybrid test has advantages over FCV and subsampling methods in most cases, but it does not dominate the SC subsampling method.

The SC methods that we consider are not asymptotically conservative, but typically are asymptotically non-similar. (That is, for tests, the limit of the supremum of the finite-sample rejection probability over points in the null hypothesis equals the nominal level, but the limit of the infimum over points in the null hypothesis is less than the nominal level.) Usually power can be improved in such cases by reducing the magnitude of asymptotic non-similarity. To do so, we introduce "plug-in" size-correction (PSC) methods for FCV, subsampling, and hybrid tests. These methods are applicable if there is a parameter sub-vector that affects the asymptotic distribution of the test statistic under consideration, is not related to the discontinuity in the asymptotic distribution, and is consistently estimable. The PSC method makes the critical values smaller for some parameter values by making the size-correction value depend on a consistent estimator of the parameter sub-vector.

The asymptotic results for subsampling methods derived in AG1, and utilized here for size correction, do not depend on the choice of subsample size b provided $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. One would expect that this may lead to poor approximations in some cases. To improve the approximations, the paper introduces finite-sample adjustments to the asymptotic rejection probabilities of subsampling and hybrid tests. The adjustments depend on the magnitude of $\delta_n = b/n$. The adjusted formulae for the asymptotic rejection probabilities are used to define adjusted SC (ASC) values and adjusted PSC (APSC) values.

All of the methods discussed above are applicable when one uses an m out of n bootstrap critical value in place of a subsampling critical value provided $m^2/n \to 0$ and the observations are i.i.d. The reason is that the m out of n bootstrap can be viewed as subsampling with replacement and the difference between sampling with and without replacement is asymptotically negligible under the stated conditions, see Politis, Romano, and Wolf (1999, p. 48).

Literature that is related to the methods considered in this paper includes the work of Politis and Romano (1994) and Politis, Romano, and Wolf (1999) on subsampling and the literature on the m out of n bootstrap, see AG1 for references. We are not aware of any methods in the literature that are analogous to the hybrid test or that consider size-correction of subsampling or m out of n bootstrap methods. Nor are we aware of any general methods of size-correction for FCV tests for the type of non-regular cases considered in this paper. For specific models in the class considered here, however, some methods are available. For example, for CIs based on post-conservative model selection estimators in regression models, Kabaila (1998) suggests a method of size-correction. For models with weak instruments, Anderson and Rubin (1949), Dufour (1997), Staiger and Stock (1997), Moreira (2003, 2009), Kleibergen (2002, 2005), Guggenberger and Smith (2005, 2008), and Otsu (2006) suggest methods with correct asymptotic size. A variant of Moreira's method also is applicable in predictive regressions with nearly integrated regressors, see Jansson and Moreira (2006). In conditionally homoskedastic autoregressive models, CI methods of Stock (1991), Andrews (1993), Nankervitz and Savin (1996), and Hansen (1999) can be used in place of the least squares estimator combined with normal critical values or subsampling critical values. Mikusheva (2007a) shows that the former methods yield correct asymptotic size. (She does not consider the method in Nankervitz and Savin (1996).)

This paper considers two examples in detail. First, we consider CIs for the autoregressive parameter ρ in a first-order conditionally heteroskedastic autoregressive (AR(1)) model in which ρ may be close to, or equal to, one. Models of this sort are applicable to exchange rate and commodity and stock prices, e.g., see Kim and Schmidt (1993). We consider FCV, subsampling, and hybrid CIs. The CIs are based on inverting a (studentized) t statistic constructed using a feasible quasi-generalized least squares (FQGLS) estimator of ρ. This is a feasible GLS estimator based on a specification of the form of the conditional heteroskedasticity that may or may not be correct. We introduce procedures that are robust to this type of misspecification. We are interested in robustness of this sort because the literature is replete with different forms of ARCH, GARCH, and stochastic volatility models for conditional heteroskedasticity—not all of which can be correct. We consider the FQGLS estimator because it has been shown that GLS correction of unit root tests yields improvements of power, see Seo (1999) and Guo and Phillips (2001).

None of the CIs in the literature, such as those in Stock (1991), Andrews (1993), Andrews and Chen (1994), Nankervis and Savin (1996), Hansen (1999), Chen and Deo (2007), and Mikusheva (2007a) have correct asymptotic size in the presence of conditional heteroskedasticity under parameter values that are not $1/n$-local-to-unity. Table 2 of Mikusheva (2007b) shows that the nominal .95 CIs of Andrews (1993), Stock (1991) (modified), and Hansen (1999) have finite-sample coverage probabilities of .70, .72, and .73, respectively, for autoregressive parameter values of .3 and .5 under conditionally normal ARCH(1) innovations with ARCH parameter equal to .85, with $n = 120$, in a model with a linear time trend. Furthermore, GLS versions of the methods listed above cannot be adapted easily to account for conditional heteroskedasticity of unknown form, which is what we consider here.[1]

[1] Stock's (1991) method is not feasible because the asymptotic distribution of the GLS t statistic

Given that the parameter space for ρ includes a unit root and near unit roots, standard two-sided FCV methods for constructing CIs based on a standard normal approximation to the t statistic are known to be problematic even under conditional homoskedasticity. As an alternative, Romano and Wolf (2001) propose subsampling CIs for ρ. Mikusheva (2007a, Theorem 4) shows that equal-tailed versions of such subsampling CIs under-cover the true value asymptotically under conditional homoskedasticity (i.e., their asymptotic confidence size is less than $1 - \alpha$), whereas some versions of the methods listed in the previous paragraph provide correct asymptotic coverage in a uniform sense. The results given here differ from those of Mikusheva (2007a) in several dimensions. First, her results do not apply to LS or FQGLS procedures in models with conditional heteroskedasticity. Second, even in a model with conditional homoskedasticity, her results do not apply to symmetric subsampling CIs and do not provide an expression for the asymptotic confidence size.

We consider two models: model 1 includes an intercept, and model 2 includes an intercept and time trend. We show that equal-tailed two-sided subsampling and two-sided FCV CIs have substantial asymptotic size distortions. On the other hand, symmetric subsampling CIs are shown to have correct asymptotic size. An explanation is given below. All types of hybrid CIs are shown to have correct asymptotic size. Finite-sample results indicate that the hybrid CIs have good coverage probabilities across all types of conditional heteroskedasticity that are considered.

The second example is a post-conservative model selection (CMS) example. We consider an LS t test concerning a regression parameter after model selection is used to determine whether another regressor should be included in the model. The model selection procedure uses an LS t test with nominal level 5%. This procedure, which is closely related to AIC, is conservative (i.e., it chooses a correct model, but not necessarily the most parsimonious model, with probability that goes to one). The asymptotic results for FCV tests in the CMS example are variations of those of Leeb (2006) and Leeb and Pötscher (2005) (and other papers referenced in these two papers).

In the CMS example, nominal 5% subsampling, FCV, and hybrid tests have asymptotic and adjusted-asymptotic sizes between 90 and 96% for upper, symmetric, and equal-tailed tests.[2] The finite-sample maximum (over the cases considered) null rejection probabilities of these tests for $n = 120$ and $b = 12$ are close to the as-

under local to unity asymptotics depends on a constant, $h_{2,7}$ below, that is unknown and is quite difficult to estimate because it depends on the unknown form of conditional heteroskedasticity which in turn depends on an infinite number of lags. The methods of Andrews (1993), Andrews and Chen (1994), and Hansen (1999) are not feasible because they depend on a parametric specification of the model since they are parametric bootstrap-type methods. Mikusheva's (2007a) procedures are variants of those considered in the papers above and hence are not easily adapted to conditional heteroskedasticity of unknown form. Chen and Deo's (2007) approach relies on the i.i.d. nature of the innovations.

[2] This is for a parameter space of $[-.995, .995]$ for the (asymptotic) correlation between the LS estimators of the two regressors.

ymptotic sizes. They are especially close for the adjusted-asymptotic sizes for which the largest deviations are 2.0%. Plug-in size-corrected tests perform very well in this example. For example, the 5% PSC hybrid test has finite-sample maximum null rejection probability of 4.8% for upper, symmetric, and equal-tailed tests.

Additional examples are given in Andrews and Guggenberger (2005, 2009b,c,d) and Guggenberger (2009). These examples cover: (i) tests when a nuisance parameter may be near a boundary of the parameter space, (ii) tests and CIs concerning the coefficient on an endogenous variable in an instrumental variables regression model with instruments that may be weak, (iii) tests concerning a parameter that determines the support of the observations, (iv) CIs constructed after the application of a consistent model selection procedure, (v) CIs when the parameter of interest may be near a boundary, (vi) tests and CIs for parameters defined by moment inequalities, and (vii) tests after the application of a pretest. Table I summarizes the asymptotic sizes of subsampling and hybrid procedures in these models for symmetric and equal-tailed two-sided procedures. In many of these models, subsampling procedures have incorrect asymptotic size—often by a substantial amount. In all of these models except those based on post-model selection, hybrid procedures have correct asymptotic size. For post-conservative model selection inference, PSC tests have correct asymptotic size.

The remainder of the paper is outlined as follows. Section 2 introduces the testing set-up, the hybrid tests, and the CMS example, which is used as a running example in the paper. Section 3 introduces the size-corrected tests, gives power comparisons of the SC tests, and introduces the plug-in size-corrected tests. Section 4 introduces the finite-sample adjustments to the asymptotic sizes of subsampling and hybrid tests. Sections 5 and 6 consider equal-tailed tests and CIs respectively. Sections 7 and 8 provide the results for the autoregressive and post-conservative model selection examples. The Supplement to the paper, Andrews and Guggenberger (2009c), gives (i) details concerning the construction of Tables II and III, (ii) proofs of the results in the paper, (iii) size-correction methods based on quantile adjustment, (iv) results concerning power comparisons of SC tests, (v) graphical illustrations of critical value functions and power comparisons, (vi) size-correction results for equal-tailed tests, (vii) results for combined size-corrected subsampling and hybrid tests, (viii) an additional example, and (ix) proofs for the examples in the paper.

2 Hybrid Tests

2.1 Intuition

We now provide some intuition regarding the potential problem with the asymptotic size of subsampling procedures and indicate why the hybrid procedure introduced below solves the problem in many cases. Suppose we are carrying out a test based on a test statistic T_n and a nuisance parameter $\gamma \in \Gamma \subset R$ appears. Suppose the asymptotic null distribution of T_n is discontinuous at $\gamma = 0$. That is, we obtain a different asymptotic distribution under the fixed parameter $\gamma = 0$ from that under a fixed $\gamma \neq 0$. As is typical in such situations, suppose the asymptotic distribution of T_n under any drifting sequence of parameters $\{\gamma_n = h/n^r : n \geq 1\}$ (or $\gamma_n = (h + o(1))/n^r$) depends on the "localization parameter" h.[3] Denote this asymptotic distribution by J_h. If the asymptotic distribution of T_n under γ_n is J_h, then the asymptotic distribution of T_b under $\gamma_n = h/n^r = (b/n)^r h/b^r = o(1)/b^r$ is J_0 when $b/n \to 0$ as $n \to \infty$. Subsample statistics with subsample size b have the same asymptotic distribution J_0 as T_b. In consequence, subsampling critical values converge in probability to the $1-\alpha$ quantile, $c_0(1-\alpha)$, of J_0, whereas the full-sample statistic T_n converges in distribution to J_h. The test statistic T_n needs a critical value equal to the $1-\alpha$ quantile, $c_h(1-\alpha)$, of J_h in order to have an asymptotic null rejection probability of α under $\{\gamma_n : n \geq 1\}$. If $c_0(1-\alpha) < c_h(1-\alpha)$, then the subsampling test over-rejects asymptotically under $\{\gamma_n : n \geq 1\}$ and has asymptotic size greater than α. If $c_0(1-\alpha) > c_h(1-\alpha)$, then it under-rejects asymptotically and is asymptotically non-similar.

Sequences of the form $\gamma_n = h/n^r$ are not the only ones in which the subsampling critical value may be too small. Suppose $\gamma_n = g/b^r$ for fixed $g \in R$ (or $\gamma_n = (g + o(1))/b^r$). Then, T_b has asymptotic distribution J_g and the probability limit of the subsampling critical value is $c_g(1-\alpha)$. On the other hand, $\gamma_n = (n/b)^r g/n^r$ and $(n/b)^r \to \infty$, so the full-sample statistic T_n converges to J_∞ (when $g \neq 0$), which is the asymptotic distribution of T_n when γ_n is more distant from the discontinuity point than $O(n^{-r})$. Let $c_\infty(1-\alpha)$ denote the $1-\alpha$ quantile of J_∞. If $c_g(1-\alpha) < c_\infty(1-\alpha)$, then the subsampling test over-rejects asymptotically under $\{\gamma_n : n \geq 1\}$. Any value of $g \in R$ is possible, so one obtains asymptotic size greater than α if $c_g(1-\alpha) < c_h(1-\alpha)$ for any (g, h) such that $g = 0$ if $h < \infty$ or $g \in R$ if $h = \infty$.[4]

The hybrid test uses a critical value given by the maximum of the subsampling critical value and $c_\infty(1-\alpha)$. Its probability limit is $c_g^* = \max\{c_g(1-\alpha), c_\infty(1-\alpha)\}$. In consequence, if the critical value function $c_h(1-\alpha)$ viewed as a function of h is maximized at $h = 0$, then when $(g, h) = (0, h)$ for $|h| < \infty$, we have $c_g^* = c_0(1-\alpha) \geq c_h(1-\alpha)$ and when (g, h) is such that $h = \infty$, we have $c_g^* \geq c_\infty(1-\alpha) = c_h(1-\alpha)$. On the other hand, if $c_h(1-\alpha)$ is maximized at $h = \infty$, then $c_g^* = c_\infty(1-\alpha) \geq c_g(1-\alpha)$ for all $g \in R \cup \{\infty\}$. Hence, in this case too the hybrid critical value does not lead to over-rejection. In many examples, $c_h(1-\alpha)$ is maximized at either 0 or ∞ and the hybrid test has correct asymptotic size.

In some models, the test statistic T_n depends on two nuisance parameters (γ_1, γ_2) and its asymptotic distribution is discontinuous whenever $\gamma_1 = 0$. In this case, the asymptotic distribution of T_n depends on a localization parameter h_1 analogous to

[3]Typically, the constant $r > 0$ is such that the distribution of T_n under γ_n is contiguous to its distribution under $\gamma = 0$. In most cases, $r = 1/2$, but in the autoregressive example with a discontinuity at a unit root, we have $r = 1$.

[4]For $g = 0$ a slightly different argument is needed.

h above and the fixed value of γ_2. The asymptotic behaviors of subsampling and hybrid tests in this case are as described above with h_1 in place of h except that the conditions for a rejection rate of $\leq \alpha$ must hold for each value of γ_2. It turns out that in a number of models of interest the critical value function is monotone increasing in h_1 for some values of γ_2 and monotone decreasing in other values. In consequence, subsampling tests over-reject asymptotically, but hybrid tests do not.

2.2 Testing Set-up

Here we describe the general testing set-up. We are interested in tests concerning a parameter $\theta \in R^d$ in the presence of a nuisance parameter $\gamma \in \Gamma$. The null hypothesis is $H_0 : \theta = \theta_0$ for some $\theta_0 \in R^d$. The alternative hypothesis may be one-sided or two-sided. Let $T_n(\theta_0)$ denote a test statistic based on a sample of size n for testing H_0. It could be a t statistic or some other test statistic. We consider the case where the asymptotic null distribution of $T_n(\theta_0)$ depends on the nuisance parameter γ and is discontinuous at some value(s) of γ.

The nuisance parameter γ has up to three components: $\gamma = (\gamma_1, \gamma_2, \gamma_3)$. The points of discontinuity of the asymptotic distribution of $T_n(\theta_0)$ are determined by the first component, $\gamma_1 \in R^p$. We assume that the discontinuities occur when one or more elements of γ_1 equal zero. The parameter space for γ_1 is $\Gamma_1 \subset R^p$. The second component, 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 $\Gamma_2 \subset R^q$. 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 $\mathcal{T}_3$. Infinite dimensional γ_3 parameters, such as error distributions, arise frequently in examples. Due to the central limit theorem (CLT), the asymptotic distribution of a test statistic often does not depend on an error distribution. The parameter space for γ_3 is $\Gamma_3(\gamma_1, \gamma_2)$ ($\subset \mathcal{T}_3$).

The parameter space for γ is

$$\Gamma = \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\}. \tag{2.1}$$

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

Assumption A. (i) Γ satisfies (2.1) and (ii) $\Gamma_1 = \prod_{m=1}^{p} \Gamma_{1,m}$, where $\Gamma_{1,m} = \lfloor \gamma_{1,m}^{\ell}, \gamma_{1,m}^{u} \rfloor$ for some $-\infty \leq \gamma_{1,m}^{\ell} < \gamma_{1,m}^{u} \leq \infty$ that satisfy $\gamma_{1,m}^{\ell} \leq 0 \leq \gamma_{1,m}^{u}$ for $m = 1, ..., p$.

Next, we describe the asymptotic behavior of $T_n(\theta_0)$ when the true value of θ is the null value θ_0. All limits are as $n \to \infty$. For an arbitrary distribution G, let $G(\cdot)$ denote the distribution function (df) of G, let $G(x-)$ denote the limit from the left of $G(\cdot)$ at x, and let $C(G)$ denote the set of continuity points of $G(\cdot)$. Let $\alpha \in (0, 1)$ be a given constant. Define the $1 - \alpha$ quantile, $q(1 - \alpha)$, of a distribution

G by $q(1-\alpha) = \inf\{x \in R : G(x) \geq 1-\alpha\}$. The distribution J_h considered below is the distribution of a proper random variable that is finite with probability one. Let $R_+ = \{x \in R : x \geq 0\}$, $R_- = \{x \in R : x \leq 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $R_{-,\infty} = R_- \cup \{-\infty\}$, $R_\infty = R \cup \{\pm\infty\}$, $R_+^p = R_+ \times ... \times R_+$ (with p copies), and $R_\infty^p = R_\infty \times ... \times R_\infty$ (with p copies).

Let $r > 0$ denote a *rate of convergence index* such that when the true parameter γ_1 satisfies $n^r \gamma_1 \to h_1$, then the test statistic $T_n(\theta_0)$ has an asymptotic distribution that depends on the localization parameter h_1. In most examples, $r = 1/2$, but in the unit root example considered below $r = 1$.

The index set for the different asymptotic null distributions of the test statistic $T_n(\theta_0)$ is

$$H = H_1 \times H_2,\ H_1 = \prod_{m=1}^{p} \begin{cases} R_{+,\infty} & \text{if } \gamma_{1,m}^\ell = 0 \\ R_{-,\infty} & \text{if } \gamma_{1,m}^u = 0 \\ R_\infty & \text{if } \gamma_{1,m}^\ell < 0 \text{ and } \gamma_{1,m}^u > 0, \end{cases} \quad \text{and } H_2 = \text{cl}(\Gamma_2), \tag{2.2}$$

where $\text{cl}(\Gamma_2)$ is the closure of Γ_2 with respect to R_∞^q. For example, if $p = 1$, $\gamma_{1,1}^\ell = 0$, and $\Gamma_2 = R^q$, then $H_1 = R_{+,\infty}$, $H_2 = R_\infty^q$, and $H = R_{+,\infty} \times R_\infty^q$. For notational simplicity, we write $h = (h_1, h_2)$, rather than $(h_1', h_2')'$, even though h is a $p + q$ column vector.

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

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

We assume that $T_n(\theta_0)$ satisfies the following conditions concerning its asymptotic null behavior.

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

Assumption K. The asymptotic distribution J_h in Assumption B is the same (proper) distribution, call it J_∞, for all $h = (h_1, h_2) \in H$ for which $h_{1,m} = +\infty$ or $-\infty$ for all $m = 1, ..., p$, where $h_1 = (h_{1,1}, ..., h_{1,p})'$.

Assumptions B and K hold in a wide variety of examples of interest, see below and Andrews and Guggenberger (2005, 2009a,c,d). In examples, when $T_n(\theta_0)$ is a studentized t statistic or a likelihood ratio (LR), Lagrange multiplier (LM), or Wald statistic, J_∞ typically is a standard normal, absolute standard normal, or chi-square distribution. Let $c_\infty(1-\alpha)$ denote the $1-\alpha$ quantile of J_∞. As defined, $c_\infty(1-\alpha)$ is an FCV that is suitable when γ is not at or close to a discontinuity point of the asymptotic distribution of $T_n(\theta_0)$.

Post-Conservative Model Selection Example. In this example, we consider inference concerning a parameter in a linear regression model after a "conservative" model selection procedure has been applied to determine whether another regressor should enter the model. A "conservative" model selection procedure is one that chooses a correct model, but not the most parsimonious correct model, with probability that goes to one as the sample size n goes to infinity. Examples are model selection based on a test whose critical value is independent of the sample size and the Akaike information criterion (AIC).

The model we consider is

$$\begin{aligned} y_i &= x_{1i}^*\theta + x_{2i}^*\beta_2 + x_{3i}^{*\prime}\beta_3 + \sigma\varepsilon_i \text{ for } i = 1, ..., n, \text{ where} \\ x_i^* &= (x_{1i}^*, x_{2i}^*, x_{3i}^{*\prime})' \in R^k, \ \beta = (\theta, \beta_2, \beta_3')' \in R^k, \end{aligned} \tag{2.3}$$

$x_{1i}^*, x_{2i}^*, \theta, \beta_2, \sigma, \varepsilon_i \in R$, and $x_{3i}^*, \beta_3 \in R^{k-2}$. The observations $\{(y_i, x_i^*) : i = 1, ..., n\}$ are i.i.d. The scaled error ε_i has mean 0 and variance 1 conditional on x_i^*.

We are interested in testing $H_0 : \theta = \theta_0$ after carrying out a model selection procedure to determine whether x_{2i}^* should enter the model. The model selection procedure is based on a t test of $H_0^* : \beta_2 = 0$ that employs a critical value c that does not depend on n. Because the asymptotic distribution of the test statistic is invariant to the value of θ_0, the testing results immediately yield results for a CI for θ obtained by inverting the test. Also, the inference problem described above covers tests concerning a linear combination of regression coefficients by suitable reparametrization (see the Supplement for details).

We consider upper and lower one-sided and symmetric and equal-tailed two-sided nominal level α FCV, subsampling, and hybrid tests of $H_0 : \theta = \theta_0$. Each test is based on a studentized test statistic $T_n(\theta_0)$, where $T_n(\theta_0)$ equals $T_n^*(\theta_0)$, $-T_n^*(\theta_0)$, $|T_n^*(\theta_0)|$, and $T_n^*(\theta_0)$, respectively, and $T_n^*(\theta_0)$ is defined below.

To define the test statistic $T_n^*(\theta_0)$, we let $\widehat{T}_{n,1}(\theta_0)$ denote the standard t statistic for testing H_0 in (2.3) (which is unrestricted in the sense that $H_0^* : \beta_2 = 0$ is not imposed). As defined, this statistic has an exact t distribution under H_0 and normality of the errors (but the latter is not assumed). We let $\widetilde{T}_{n,1}(\theta_0)$ denote the "restricted" t statistic for testing H_0 which imposes the restriction of $H_0^* : \beta_2 = 0$, but uses the unrestricted estimator $\widehat{\sigma}$ of σ instead of the restricted estimator.[5] We let $T_{n,2}$ denote the standard t statistic for testing $H_0^* : \beta_2 = 0$ (and does not impose H_0).[6] The model selection test rejects $H_0^* : \beta_2 = 0$ if $|T_{n,2}| > c$, where $c > 0$ is a given critical value

[5] One could define $\widetilde{T}_{n,1}(\theta_0)$ using the restricted (by $\beta_2 = 0$) LS estimator of σ, but this is not desirable because it leads to an inconsistent estimator of σ under sequences of parameters $\{\beta_2 = \beta_{2n} : n \geq 1\}$ that satisfy $\beta_{2n} \to 0$ and $n^{1/2}\beta_{2n} \nrightarrow 0$ as $n \to \infty$. For subsampling tests, one could define $\widetilde{T}_{n,1}(\theta_0)$ and $\widehat{T}_{n,1}(\theta_0)$ with $\widehat{\sigma}$ deleted because the scale of the subsample statistics offsets that of the original sample statistic. This does not work for hybrid tests because Assumption K fails if $\widehat{\sigma}$ is deleted.

[6] See Section 11.1 of the Supplement for explicit expressions for $\widehat{T}_{n,1}(\theta_0)$, $\widetilde{T}_{n,1}(\theta_0)$, and $T_{n,2}$.

that does not depend on n. Typically, $c = z_{1-\alpha/2}$ for some $\alpha > 0$. The post-model selection test statistic, $T_n^*(\theta_0)$, for testing $H_0 : \theta = \theta_0$ is

$$T_n^*(\theta_0) = \widetilde{T}_{n,1}(\theta_0) 1(|T_{n,2}| \le c) + \widehat{T}_{n,1}(\theta_0) 1(|T_{n,2}| > c). \tag{2.4}$$

We now show how the testing problem above fits into the general framework. First, we define the regressor vector $x_i^\perp = (x_{1i}^\perp, x_{2i}^\perp)'$ that corresponds to $(x_{1i}^*, x_{2i}^*)'$ with x_{3i}^* projected out using the population projection. Let G denote the distribution of (ε_i, x_i^*). Define

$$\begin{aligned} x_i^\perp &= \begin{pmatrix} x_{1i}^* - x_{3i}^{*\prime}(E_G x_{3i}^* x_{3i}^{*\prime})^{-1} E_G x_{3i}^* x_{1i}^* \\ x_{2i}^* - x_{3i}^{*\prime}(E_G x_{3i}^* x_{3i}^{*\prime})^{-1} E_G x_{3i}^* x_{2i}^* \end{pmatrix} \in R^2, \\ Q &= E_G x_i^\perp x_i^{\perp\prime}, \text{ and } Q^{-1} = \begin{bmatrix} Q^{11} & Q^{12} \\ Q^{12} & Q^{22} \end{bmatrix}. \end{aligned} \tag{2.5}$$

The parameter vector $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ is defined in this example by

$$\gamma_1 = \frac{\beta_2}{\sigma (Q^{22})^{1/2}}, \ \gamma_2 = \frac{Q^{12}}{(Q^{11} Q^{22})^{1/2}}, \text{ and } \gamma_3 = (\beta_2, \beta_3, \sigma, G). \tag{2.6}$$

Note that $\gamma_2 = \rho$, where $\rho = AsyCorr(\widehat{\theta}, \widehat{\beta}_2)$. The parameter spaces for γ_1, γ_2, and γ_3 are $\Gamma_1 = R$, $\Gamma_2 = [-1+\zeta, 1-\zeta]$ for some $\zeta > 0$, and

$$\begin{aligned} \Gamma_3(\gamma_1, \gamma_2) = \Big\{ &(\beta_2, \beta_3, \sigma, G) : \beta_2 \in R, \ \beta_3 \in R^{k-2}, \ \sigma > 0, \text{ and for} \\ &Q = E_G x_i^\perp x_i^{\perp\prime} \text{ and } Q^{-1} = \begin{bmatrix} Q^{11} & Q^{12} \\ Q^{12} & Q^{22} \end{bmatrix}, \text{ (i) } \frac{\beta_2}{\sigma (Q^{22})^{1/2}} = \gamma_1, \\ &\text{(ii) } \frac{Q^{12}}{(Q^{11} Q^{22})^{1/2}} = \gamma_2, \text{ (iii) } \lambda_{\min}(Q) \ge \kappa, \text{ (iv) } \lambda_{\min}(E_G x_{3i}^* x_{3i}^{*\prime}) \ge \kappa, \\ &\text{(v) } E_G ||x_i^*||^{2+\delta} \le M, \text{ (vi) } E_G ||\varepsilon_i x_i^*||^{2+\delta} \le M, \\ &\text{(vii) } E_G(\varepsilon_i | x_i^*) = 0 \text{ a.s., and (viii) } E_G(\varepsilon_i^2 | x_i^*) = 1 \text{ a.s.} \Big\} \end{aligned} \tag{2.7}$$

for some $\kappa, \delta > 0$ and $M < \infty$. The parameter γ_2 is bounded away from one and minus one because otherwise the LS estimators of θ and β_2 could have a distribution that is arbitrarily close to being singular (such as a normal distribution with singular variance matrix). Assumption A holds immediately.

The rate of convergence parameter r equals $1/2$. The localization parameter h satisfies $h = (h_1, h_2) \in H = H_1 \times H_2$, where $H_1 = R_\infty$ and $H_2 = [-1+\zeta, 1-\zeta]$.

Let $\Delta(a, b) = \Phi(a+b) - \Phi(a-b)$, where $\Phi(\cdot)$ is the standard normal distribution function. Note that $\Delta(a, b) = \Delta(-a, b)$. Calculations in the Supplement to this paper establish that the asymptotic distribution J_h^* of $T_n^*(\theta_0)$ under a sequence of parameters $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) : n \ge 1\}$ (where $n^{1/2}\gamma_{n,1} \to h_1$, $\gamma_{n,2} \to h_2$, and $\gamma_{n,3} \in \Gamma_3(\gamma_{n,1}, \gamma_{n,2})$ for all n) is

$$\begin{aligned} J_h^*(x) &= \Phi(x + h_1 h_2 (1 - h_2^2)^{-1/2}) \Delta(h_1, c) \\ &\quad + \int_{-\infty}^x \left(1 - \Delta\left(\frac{h_1 + h_2 t}{(1-h_2^2)^{1/2}}, \frac{c}{(1-h_2^2)^{1/2}} \right) \right) \phi(t) dt \end{aligned} \tag{2.8}$$

when $|h_1| < \infty$. When $|h_1| = \infty$, $J_h^*(x) = \Phi(x)$ (which equals the limit as $|h_1| \to \infty$ of $J_h^*(x)$ defined in (2.8)). For upper one-sided, lower one-sided, and symmetric two-sided tests, the asymptotic distribution J_h of $T_n(\theta_0)$ is given by J_h^*, $-J_h^*$, and $|J_h^*|$, respectively. (If $Y \sim J_h^*$, then by definition, $-Y \sim -J_h^*$ and $|Y| \sim |J_h^*|$.) This verifies Assumption B. Assumption K holds with J_∞^* being a $N(0,1)$ distribution.

The asymptotic results that are used to verify Assumption B are closely related to results of Leeb (2006) and Leeb and Pötscher (2005) (and other papers referenced in these two papers). However, no papers in the literature, that we are aware of, consider subsampling-based methods for post-conservative model selection procedures, as is done below. The results given below also are related to, but quite different from, those in Andrews and Guggenberger (2009b) for post-*consistent* model selection estimators, shrinkage estimators, and super-efficient estimators.

2.3 Subsampling Critical Value

The hybrid test introduced below makes use of a subsampling critical value, which we define here. A subsampling critical value is determined by subsample statistics that are denoted by $\{\widehat{T}_{n,b,j} : j = 1, ..., q_n\}$, where j indexes the subsample, b is a subsample size that depends on n, and q_n is the number of different subsamples. With i.i.d. observations, there are $q_n = n!/((n-b)!b!)$ different subsamples of size b. With time series observations, there are $q_n = n - b + 1$ subsamples each consisting of b consecutive observations.

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 are based on subsamples of size b rather than the full sample. The subsample statistics $\{\widehat{T}_{n,b,j} : j = 1, ..., q_n\}$ that are used to construct the subsampling critical value are defined to satisfy one or the other of the following assumptions.

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.

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

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

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

$$L_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(\widehat{T}_{n,b,j} \leq x) \text{ for } x \in R \text{ and}$$
$$c_{n,b}(1-\alpha) = \inf\{x \in R : L_{n,b}(x) \geq 1-\alpha\}. \quad (2.9)$$

The subsampling critical value is $c_{n,b}(1-\alpha)$. The subsampling test rejects $H_0 : \theta = \theta_0$ if $T_n(\theta_0) > c_{n,b}(1-\alpha)$.

For subsampling tests (and the hybrid tests introduced below), we assume:

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

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

These assumptions allow for i.i.d., stationary strong-mixing, and even nonstationary observations (as shown in the autoregressive example below). They have been verified in a wide variety of examples in this paper and elsewhere.

In the post-conservative model selection example, the subsampling critical values are defined using Assumption Sub1. Let $\widetilde{\theta}$ and $\widehat{\theta}$ denote the restricted and unrestricted least squares (LS) estimators of θ, respectively. The subsample statistics are defined by $\{T_{n,b,j}(\overline{\theta}) : j = 1, ..., q_n\}$, where $\overline{\theta}$ is the "model-selection" estimator of θ defined by

$$\overline{\theta} = \widetilde{\theta}1(|T_{n,2}| \leq c) + \widehat{\theta}1(|T_{n,2}| > c) \tag{2.10}$$

and $T_{n,b,j}(\theta_0)$ is defined just as $T_n(\theta_0)$ is defined but using the jth subsample of size b in place of the full sample of size n. (One could also use the unrestricted estimator $\widehat{\theta}$ in place of $\overline{\theta}$.) Assumption C holds by choice of b. Assumption D holds automatically.

2.4 Technical Assumptions

We now state several technical assumptions that are used below. Define the empirical distribution of $\{T_{n,b,j}(\theta_0) : j = 1, ..., q_n\}$ by $U_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta_0) \leq x)$.

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

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

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

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

Assumption E holds for i.i.d. observations and for stationary strong-mixing observations with $\sup_{\gamma \in \Gamma} \alpha_\gamma(m) \to 0$ as $m \to \infty$, where $\{\alpha_\gamma(m) : m \geq 1\}$ are the strong-mixing numbers of the observations when the true parameters are (θ_0, γ), see AG1. Assumptions F and J are not very restrictive. The former is used for one- and two-sided tests, while the latter is used for equal-tailed tests. Assumption G holds automatically when $\{\widehat{T}_{n,b,j}\}$ satisfy Assumption Sub2. Section 7 of AG1 provides sufficient conditions for Assumption G when Assumption Sub1 holds.

In the post-conservative model selection example, Assumption E holds automatically, Assumptions F and J hold because $J_h^*(x)$ is strictly increasing in $x \in R$ for all $h \in H$, and Assumption G is verified in the Supplement to this paper using the proof of Lemma 4 of AG1.

2.5 Definition of Hybrid Tests

We now introduce a hybrid test that is useful when the test statistic $T_n(\theta_0)$ has a limit distribution that is discontinuous in some parameter and an FCV or subsampling test over-rejects asymptotically under the null hypothesis. The critical value of the hybrid test is the maximum of the subsampling critical value and a certain fixed critical value. The hybrid test is quite simple to compute, in many situations has asymptotic size equal to its nominal level α, see Lemma 2 below and the examples in Table I, and otherwise over-rejects the null asymptotically less than the standard subsampling test or the FCV test at some null parameter values. In addition, in many scenarios, the power of the hybrid test is quite good relative to FCV and subsampling tests (after all have been size-corrected), see Section 3.2 below.

The hybrid test with nominal level α rejects the null hypothesis $H_0 : \theta = \theta_0$ when

$$\begin{aligned} &T_n(\theta_0) > c^*_{n,b}(1-\alpha), \text{ where} \\ &c^*_{n,b}(1-\alpha) = \max\{c_{n,b}(1-\alpha), c_\infty(1-\alpha)\}. \end{aligned} \tag{2.11}$$

The hybrid test simply takes the critical value to be the maximum of the usual subsampling critical value and the critical value from the J_∞ distribution, which is usually known.[7] For example, in the post-conservative model selection example, $c_\infty(1-\alpha)$ equals $z_{1-\alpha}$ and $z_{1-\alpha/2}$ for one- and two-sided tests, respectively. Hence, the hybrid test is straightforward to compute. Obviously, the rejection probability of the hybrid test is less than or equal to those of the standard subsampling test and the FCV test with critical value $c_\infty(1-\alpha)$. Hence, the hybrid test does not over-reject more often than both of these two tests.

Furthermore, it is shown in Lemma 2 below that the hybrid test of nominal level α has asymptotic size α provided the $1-\alpha$ quantile function $c_{(h_1,h_2)}(1-\alpha)$ of $J_{(h_1,h_2)}$ is maximized at a boundary point of h_1 for each fixed h_2, where $h = (h_1, h_2)$. For example, this occurs if $c_h(1-\alpha)$ is monotone increasing or decreasing in h_1 for each fixed $h_2 \in H_2$.

[7] Hybrid tests can be defined even when Assumption K does not hold. For example, we can define $c^*_{n,b}(1-\alpha) = \max\{c_{n,b}(1-\alpha), \sup_{h\in H} c_{h^\infty}(1-\alpha)\}$, where $c_{h^\infty}(1-\alpha)$ is the $1-\alpha$ quantile of J_{h^∞} and, given $h = (h_1, h_2) \in H$, $h^\infty = (h^\infty_{1,1}, ..., h^\infty_{1,p}, h^\infty_2) \in H$ is defined by $h^\infty_{1,j} = +\infty$ if $h_{1,j} > 0$, $h^\infty_{1,j} = -\infty$ if $h_{1,j} < 0$, $h^\infty_{1,j} = +\infty$ or $-\infty$ (chosen so that $h^\infty \in H$) if $h_{1,j} = 0$ for $j = 1, ..., p$, and $h^\infty_2 = h_2$. When Assumption K holds, this reduces to the hybrid critical value in (2.11). We utilize Assumption K because it leads to a particularly simple form for the hybrid test.

2.6 Asymptotic Size

The exact and asymptotic size of a hybrid test are:

$$\begin{aligned} ExSz_n(\theta_0) &= \sup_{\gamma \in \Gamma} P_{\theta_0,\gamma}(T_n(\theta_0) > c^*_{n,b}(1-\alpha)) \text{ and} \\ AsySz(\theta_0) &= \limsup_{n\to\infty} ExSz_n(\theta_0), \end{aligned} \tag{2.12}$$

where $P_{\theta,\gamma}(\cdot)$ denotes probability when the true parameters are (θ,γ). We are interested in the "asymptotic size" of the test because it approximates the exact size. Uniformity over $\gamma \in \Gamma$, which is built into the definition of asymptotic size, is necessary for asymptotic results to give a good approximation to the exact size.

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

$$\begin{aligned} GH = \{&(g,h) \in H \times H : g = (g_1, g_2),\ h = (h_1, h_2),\ g_2 = h_2 \text{ and for } m = 1, ..., p, \\ &\text{(i) } g_{1,m} = 0 \text{ if } |h_{1,m}| < \infty,\ \text{(ii) } g_{1,m} \in R_{+,\infty} \text{ if } h_{1,m} = +\infty, \text{ and} \\ &\text{(iii) } g_{1,m} \in R_{-,\infty} \text{ if } h_{1,m} = -\infty\}, \end{aligned} \tag{2.13}$$

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 (defined as the case where there is no γ_1 component of γ), GH simplifies considerably: $GH = \{(g_2, h_2) \in H_2 \times H_2 : g_2 = h_2\}$. See AG1 for further discussion of GH.

Define

$$Max_{Hyb}(\alpha) = \sup_{(g,h)\in GH} [1 - J_h(\max\{c_g(1-\alpha), c_\infty(1-\alpha)\})]. \tag{2.14}$$

If $J_h(x)$ is continuous at suitable (h,x) values, then the following assumption holds.

Assumption T. $Max_{Hyb}(\alpha) = Max^-_{Hyb}(\alpha)$, where $Max^-_{Hyb}(\alpha)$ is defined as $Max_{Hyb}(\alpha)$ is defined in (2.14), but with $J_h(x)$ replaced by $J_h(x-)$, where $x = \max\{c_g(1-\alpha), c_\infty(1-\alpha)\}$.

Assumption T holds in the post-conservative model selection example by the continuity of $J^*_h(x)$ in x for $x \in R$ for all $h \in H$. It also holds in all of the examples we have considered except the moment inequality example.[8]

The following result establishes the asymptotic size of the hybrid test.

[8] Assumption T is not needed in the moment inequality example because subsampling has correct asymptotic size in that example, see Andrews and Guggenberger (2009d).

Theorem 1 *Suppose Assumptions* A-G, K, *and* T *hold. Then, the hybrid test based on* $T_n(\theta_0)$ *has* $AsySz(\theta_0) = Max_{Hyb}(\alpha)$.

Comment. Theorem 1 holds by the proof of Theorem 1(ii) of AG1 with $c_{n,b}(1-\alpha)$ replaced by $\max\{c_{n,b}(1-\alpha), c_\infty(1-\alpha)\}$ throughout using a slight variation of Lemma 5(ii) of AG1.

2.7 Properties of Hybrid Tests

The following result shows that the hybrid test has better size properties than the subsampling test. It is shown in AG1 that the subsampling test has asymptotic size that satisfies $AsySz(\theta_0) = Max_{Sub}(\alpha)$, where $Max_{Sub}(\alpha)$ is defined just as $Max_{Hyb}(\alpha)$ is, but with $c_g(1-\alpha))$ in place of $\max\{c_g(1-\alpha), c_\infty(1-\alpha)\}$.

Lemma 1 *Suppose Assumptions* A-G, K, *and* T *hold. Then, either* (i) *the addition of* $c_\infty(1-\alpha)$ *to the subsampling critical value is irrelevant asymptotically (i.e.,* $c_h(1-\alpha) \geq c_\infty(1-\alpha)$ *for all* $h \in H$ *and* $Max_{Hyb}(\alpha) = Max_{Sub}(\alpha)$*), or* (ii) *the nominal level* α *subsampling test over-rejects asymptotically (i.e.,* $AsySz(\theta_0) > \alpha$*) and the hybrid test reduces the asymptotic over-rejection for at least one parameter value* $(g, h) \in GH$.

Next, we show that the hybrid test has correct size asymptotically if $c_h(1-\alpha)$ is maximized at h^∞ or is maximized at $h^0 = (0, h_2)$ and $p = 1$, where p is the dimension of h_1 and h^∞ is any $h \in H$ for which $J_h = J_\infty$. For example, for $p = 1$, the maximization condition is satisfied if $c_h(1-\alpha)$ is monotone increasing or decreasing in h_1, is bowl-shaped in h_1, or is wiggly in h_1 with global maximum at 0 or $\pm\infty$. The precise condition is the following. (Here, "Quant" abbreviates "quantile.")

Assumption Quant. (i) (a) for all $h \in H$, $c_h(1-\alpha) \leq c_\infty(1-\alpha)$ and (b) $J_\infty(c_\infty(1-\alpha)) = 1-\alpha$; or (ii) (a) $p = 1$, (b) for all $h \in H$, $c_h(1-\alpha) \leq c_{h^0}(1-\alpha)$, and (c) $J_\infty(c_\infty(1-\alpha)) = 1-\alpha$.

Assumption Quant (i)(b) and (ii)(c) are continuity conditions that are not restrictive.

Lemma 2 *Suppose Assumptions* A-G, K, T, *and* Quant *hold. Then, the hybrid test based on* $T_n(\theta_0)$ *has* $AsySz(\theta_0) = \alpha$.

3 Size-Corrected Tests

3.1 Definition and Justification of Size-Corrected Tests

We now define size-corrected (SC) tests. The size-corrected fixed critical value (SC-FCV), subsampling (SC-Sub), and hybrid (SC-Hyb) tests with nominal level α

are defined to reject the null hypothesis $H_0 : \theta = \theta_0$ when

$$\begin{aligned} T_n(\theta_0) &> cv(1-\alpha), \\ T_n(\theta_0) &> c_{n,b}(1-\alpha) + \kappa(\alpha) \text{ and} \\ T_n(\theta_0) &> \max\{c_{n,b}(1-\alpha), c_\infty(1-\alpha) + \kappa^*(\alpha)\}, \end{aligned} \tag{3.1}$$

respectively, where

$$\begin{aligned} cv(1-\alpha) &= \sup_{h \in H} c_h(1-\alpha), \\ \kappa(\alpha) &= \sup_{(g,h) \in GH} [c_h(1-\alpha) - c_g(1-\alpha)], \\ \kappa^*(\alpha) &= \sup_{h \in H^*} c_h(1-\alpha) - c_\infty(1-\alpha), \text{ and} \\ H^* &= \{h \in H : \text{for some } (g,h) \in GH, c_g(1-\alpha) < c_h(1-\alpha)\}. \end{aligned} \tag{3.2}$$

If H^* is empty, then $\kappa^*(\alpha) = -\infty$ by definition.

Size correction as in (3.1) is possible under the following assumption.

Assumption L. (i) $\sup_{h \in H} c_h(1-\alpha) < \infty$ and (ii) $\inf_{h \in H} c_h(1-\alpha) > -\infty$.

Assumption L is satisfied in most, but not all, examples. Assumption L holds in the post-conservative model selection example because $c_h(1-\alpha)$ is continuous in $h \in H$ and has finite limits as $|h_1| \to \infty$ and/or $|h_2| \to 1 - \zeta$. Assumption L(i) is a necessary and sufficient condition for size correction of the FCV test. Necessary and sufficient conditions for size correction of the subsampling and hybrid tests are given in Andrews and Guggenberger (2005). These conditions are weaker than Assumption L, but more complicated. Even the weaker conditions are violated in some examples, e.g., in the consistent model selection/super-efficient example in Andrews and Guggenberger (2009b).

In some cases the FCV test cannot be size-corrected because $cv(1-\alpha) = \infty$, but the SC-Sub and SC-Hyb tests still exist and have correct asymptotic size. Also, in some cases, the SC-FCV and SC-Hyb tests exist while the SC-Sub test does not (because $\kappa(\alpha) = \infty$). Surprisingly, both cases arise in the instrumental variables (IV) example considered in Andrews and Guggenberger (2005) (depending upon whether one considers symmetric two-sided or upper one-sided tests).

The following is a continuity condition that is not very restrictive.

Assumption M. (a) (i) For some $h^* \in H$, $c_{h^*}(1-\alpha) = \sup_{h \in H} c_h(1-\alpha)$ and (ii) for all $h^* \in H$ that satisfy the condition in part (i), $J_{h^*}(x)$ is continuous at $x = c_{h^*}(1-\alpha)$.
(b) (i) For some $(g^*, h^*) \in GH$, $c_{h^*}(1-\alpha) - c_{g^*}(1-\alpha) = \sup_{(g,h) \in GH} [c_h(1-\alpha) - c_g(1-\alpha)]$ and (ii) for all $(g^*, h^*) \in GH$ that satisfy the condition in part (i), $J_{h^*}(x)$ is continuous at $x = c_{h^*}(1-\alpha)$.
(c) (i) When H^* is non-empty, for some $h^* \in H^*$, $c_{h^*}(1-\alpha) = \sup_{h \in H^*} c_h(1-\alpha)$ and (ii) for all $(g,h) \in GH$ with $\max\{c_g(1-\alpha), c_\infty(1-\alpha) + \kappa^*(\alpha)\} = c_h(1-\alpha)$, $J_h(x)$ is continuous at $x = c_h(1-\alpha)$.

Assumption M holds in the post-conservative model selection example by the continuity of $c_h(1-\alpha)$ in $h \in H$ plus the shape of $c_h(1-\alpha)$ as a function of h_1 for each $|h_2| \leq 1-\zeta$ (which is determined by simulation), see Figure 2 below.

The following result shows that the SC tests have $AsySz(\theta_0)$ equal to their nominal level under suitable assumptions.

Theorem 2 *Suppose Assumptions* A-G *and* K-M *hold. Then, the* SC-FCV, SC-Sub, *and* SC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.

Comments. 1. The proof of Theorem 2 can be altered slightly to prove that $\lim_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}(T_n(\theta_0) > cv(1-\alpha)) = \alpha$ for the SC-FCV test under the given assumptions (which is slightly stronger than the result in Theorem 2) and analogously for the SC-Sub and SC-Hyb tests.

2. Assumptions C-G are only used for the SC-Sub and SC-Hyb tests. Assumption K is only used for the SC-Hyb test. Part (a) of Assumption M is only used for the SC-FCV test and analogously part (b) only for the SC-Sub test and part (c) only for SC-Hyb test.

To compute $cv(1-\alpha), \kappa(\alpha)$, and $\kappa^*(\alpha)$, one needs to be able to compute $c_h(1-\alpha)$ for $h \in H$ and carry out maximization over $h \in H$ or $(g,h) \in GH$. Computation of $c_h(1-\alpha)$ can be done analytically in some cases, by numerical integration if the density of J_h is available, or by simulation. The maximization step may range in difficulty from being very easy to nearly impossible depending on the dimension $p+q$ of h, the shape and smoothness of $c_h(1-\alpha)$ as a function of h, and the time needed to compute $c_h(1-\alpha)$ for any given h. For a given example, one can tabulate $cv(1-\alpha)$, $\kappa(\alpha)$, and $\kappa^*(\alpha)$ for selected values of α. Once this is done, the SC-FCV, SC-Sub, and SC-Hyb tests are as easy to apply as the corresponding non-corrected tests.

An alternative method of size-correcting subsampling and hybrid tests is to adjust the quantile of the test rather than to increase the critical value by a fixed amount, see the Supplement.

3.2 Power Comparisons of Size-Corrected Tests

Here we compare the asymptotic power of the SC-FCV, SC-Sub, and SC-Hyb tests. Since all three tests employ the same test statistic $T_n(\theta_0)$, the comparison is based on the magnitudes of the critical values of the tests for n large. The SC-FCV critical value is fixed. The other two critical values are random and their large sample behavior depends on the sequence $\{\gamma_n \in \Gamma : n \geq 1\}$ of true parameters. We focus on the case in which these critical values do not depend on whether the null hypothesis is true, which typically holds when the subsample statistics are defined to satisfy Assumption Sub1 (and fails when they satisfy Assumption Sub2).

The possible limits of the SC-Sub and SC-Hyb critical values under $\{\gamma_{n,h}\}$ are

$$c_g(1-\alpha) + \kappa(\alpha) \ \& \ \max\{c_g(1-\alpha), c_\infty(1-\alpha) + \kappa^*(\alpha)\} \text{ for } g \in H \tag{3.3}$$

(see Lemma 6(v) of AG1). The relative magnitudes of the limits of the critical values are determined by the shapes of the quantiles $c_g(1-\alpha)$ as functions of $g \in H$.

The first result is that the SC-Hyb test is always at least as powerful as the SC-FCV test. This holds because for all $g \in H$,

$$\begin{aligned} \max\{c_g(1-\alpha), c_\infty(1-\alpha) + \kappa^*(\alpha)\} &= \max\{c_g(1-\alpha), \sup_{h \in H^*} c_h(1-\alpha)\} \\ &\leq \sup_{h \in H} c_h(1-\alpha) = cv(1-\alpha). \end{aligned} \tag{3.4}$$

The same is not true of the SC-Sub test vis-a-vis the SC-FCV test.

Next, Theorem S1 in the Supplement shows that (a) if $c_g(1-\alpha) \geq c_h(1-\alpha)$ for all $(g,h) \in GH$, then the SC-Sub, SC-Hyb, Sub, and Hyb tests are equivalent asymptotically and are more powerful than the SC-FCV test; (b) if $c_g(1-\alpha) \leq c_h(1-\alpha)$ for all $(g,h) \in GH$, then the SC-FCV, SC-Hyb, FCV, and Hyb tests are equivalent asymptotically and are more powerful than the SC-Sub test; and (c) if $H = H_1 = R_{+,\infty}$ and $c_h(1-\alpha)$ is uniquely maximized at $h^* \in (0,\infty)$, then the SC-FCV and SC-Hyb tests are asymptotically equivalent and are either (i) more powerful than the SC-Sub test for all $(g,h) \in GH$, or (ii) more powerful than the SC-Sub test for some values of $(g,h) \in GH$ but less powerful for other values of $(g,h) \in GH$. Note that these power comparisons hold even if different subsample sizes are used for the hybrid and subsampling procedures provided both satisfy $b \to \infty$ and $b/n \to 0$ (because the asymptotic results do not depend on the specific choice of b).

These results show that the SC-Hyb test has some nice power properties. When the SC-Sub test dominates the SC-FCV test, the SC-Hyb test behaves like the SC-Sub test. When the SC-FCV test dominates the SC-Sub test, the SC-Hyb test behaves like the SC-FCV test. In none of the cases considered is the SC-Hyb test dominated by the SC-FCV or SC-Sub tests.

3.3 Plug-in Size-Corrected Tests

Here, we introduce improved size-correction methods that employ a consistent estimator $\widehat{\gamma}_{n,2}$ of the nuisance parameter γ_2. The idea is to size correct a test differently for different values of $\widehat{\gamma}_{n,2}$, rather than size correct by a value that is sufficiently large to work uniformly for all $\gamma_2 \in \Gamma_2$. This yields a more powerful test. The estimator $\widehat{\gamma}_{n,2}$ is assumed to satisfy the following assumption.

Assumption N. $\widehat{\gamma}_{n,2} - \gamma_{n,2} \to_p 0$ under all sequences $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\}$.

Assumption N holds in many cases. But it fails in models that are unidentified at the discontinuity point of the asymptotic distribution of $T_n(\theta_0)$, as occurs in an IV regression model with IVs that may be weak.

Define

$$cv_{h_2}(1-\alpha) = \sup_{h_1 \in H_1} c_{(h_1,h_2)}(1-\alpha),$$

$$\kappa_{h_2}(\alpha) = \sup_{g_1,h_1\in H_1:((g_1,h_2),(h_1,h_2))\in GH} (c_{(h_1,h_2)}(1-\alpha) - c_{(g_1,h_2)}(1-\alpha)), \text{ and}$$
$$\kappa^*_{h_2}(\alpha) = \sup_{h_1\in H^*_{h_2}} c_{(h_1,h_2)}(1-\alpha) - c_\infty(1-\alpha), \text{ where}$$
$$H^*_{h_2} = \{h_1 \in H_1 : \text{for some } g_1 \in H_1,\ (g,h) = ((g_1,h_2),(h_1,h_2)) \in GH, \ \& \ c_g(1-\alpha) < c_h(1-\alpha)\}. \quad (3.5)$$

If $H^*_{h_2}$ is empty, then $\kappa^*_{h_2}(\alpha) = -\infty$. The PSC-FCV, PSC-Sub, and PSC-Hyb tests are defined as in (3.1) with $cv(1-\alpha)$, $\kappa(\alpha)$, and $\kappa^*(\alpha)$ replaced by $cv_{\widehat{\gamma}_{n,2}}(1-\alpha)$, $\kappa_{\widehat{\gamma}_{n,2}}(\alpha)$, and $\kappa^*_{\widehat{\gamma}_{n,2}}(\alpha)$, respectively.

Clearly, $cv_{\widehat{\gamma}_{n,2}}(1-\alpha) \le cv(1-\alpha)$ (with strict inequality whenever $\widehat{\gamma}_{n,2}$ takes a value that does not maximize $cv_{h_2}(1-\alpha)$ over $h_2 \in H_2$). In consequence, the PSC-FCV test is asymptotically more powerful than the SC-FCV test. Analogous results hold for the critical values and asymptotic power of the PSC-Sub and PSC-Hyb tests relative to the SC-Sub and SC-Hyb tests.

The following continuity assumption is not very restrictive.

Assumption O. (a) (i) $cv_{h_2}(1-\alpha)$ is uniformly continuous in h_2 on H_2, (ii) for each $h_2 \in H_2$, there exists some $h_1^* \in H_1$ such that $c_{(h_1^*,h_2)}(1-\alpha) = cv_{h_2}(1-\alpha)$, and (iii) for all $h = (h_1,h_2) \in H$ for which $c_h(1-\alpha) = cv_{h_2}(1-\alpha)$, $J_h(x)$ is continuous at $x = cv_{h_2}(1-\alpha)$.
(b) (i) $\kappa_{h_2}(\alpha)$ is uniformly continuous in h_2 on H_2, (ii) for each $h_2 \in H_2$, there exists some $g_1^*, h_1^* \in H_1$ such that $(g^*,h^*) = ((g_1^*,h_2),(h_1^*,h_2)) \in GH$ and $c_{(h_1^*,h_2)}(1-\alpha) - c_{(g_1^*,h_2)}(1-\alpha) = \kappa_{h_2}(1-\alpha)$, and (iii) for all $(g,h) \in GH$ for which $c_h(1-\alpha) - c_g(1-\alpha) = \kappa_{h_2}(1-\alpha)$, where $h = (h_1,h_2)$, $J_h(x)$ is continuous at $x = c_g(1-\alpha) + \kappa_{h_2}(1-\alpha)$.
(c) (i) $\kappa^*_{h_2}(\alpha)$ is uniformly continuous in h_2 on H_2, (ii) for each $h_2 \in H_2$, when $H^*_{h_2}$ is non-empty, we have: for some $h_1^* \in H^*_{h_2}$, $c_{(h^*_1,h_2)}(1-\alpha) - c_\infty(1-\alpha) = \kappa^*_{h_2}(1-\alpha)$, and (iii) for all $(g,h) \in GH$ for which $c_h(1-\alpha) - c_\infty(1-\alpha) = \kappa^*_{h_2}(1-\alpha)$, where $h = (h_1,h_2)$, $J_h(x)$ is continuous at $x = \max\{c_g(1-\alpha), c_\infty(1-\alpha) + \kappa^*_{h_2}(1-\alpha)\}$.

Assumption O holds in the post-conservative model selection example given the definition of $J^*_h(x)$ in (2.8).

Theorem 3 *Suppose Assumptions* A-G, K, L, N, *and* O *hold. Then,* (a) $cv_{\widehat{\gamma}_{n,2}}(1-\alpha) - cv_{\gamma_{n,2}}(1-\alpha) \to_p 0$, $\kappa_{\widehat{\gamma}_{n,2}}(\alpha) - \kappa_{\gamma_{n,2}}(\alpha) \to_p 0$, *and* $\kappa^*_{\widehat{\gamma}_{n,2}}(\alpha) - \kappa^*_{\gamma_{n,2}}(\alpha) \to_p 0$ *under all sequences* $\{\gamma_n = (\gamma_{n,1},\gamma_{n,2},\gamma_{n,3}) \in \Gamma : n \ge 1\}$, *and* (b) *the* PSC-FCV, PSC-Sub, *and* PSC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.

Comment. Assumption O(a) is only used for the PSC-FCV test and likewise part (b) is only used for the PSC-Sub test and part (c) for the PSC-Hyb test.

4 Finite-Sample Adjustments

In this section, we introduce a finite-sample adjustment to the $AsySz(\theta_0)$ of subsampling and hybrid tests. It is designed to give a better approximation to the actual finite-sample sizes of these tests than does $AsySz(\theta_0)$. The adjustments are used to construct finite-sample adjusted size-corrected (ASC) subsampling and hybrid tests, both with and without plug-in estimation of h_2. The idea of the adjustment is to retain the actual ratio $\delta_n = b/n$ of the subsample size to the full-sample size in the approximation to the exact size of the tests, rather than to use its asymptotic limit, which is zero.

The adjustment method is described roughly as follows. For simplicity, consider the case in which γ does not contain subvectors γ_2 or γ_3, $p = 1$, and $\Gamma = [0, d]$ for some $0 < d < \infty$. Under Assumption B, the distribution of $T_n(\theta_0)$ under γ can be approximated by J_{h_n}, where $h_n = n^r\gamma$. Hence, the distribution of $T_b(\theta_0)$ under γ can be approximated by $J_{h_n^*}$, where $h_n^* = b^r\gamma = (b/n)^r h_n = \delta_n^r h_n$. In turn, the $1-\alpha$ subsampling quantile $c_{n,b}(1-\alpha)$ under γ can be approximated by the $1-\alpha$ quantile of $J_{h_n^*} = J_{\delta_n^r h_n}$, viz., $c_{\delta_n^r h_n}(1-\alpha)$. This leads to the approximation of $P_{\theta_0,\gamma}(T_n(\theta_0) > c_{n,b}(1-\alpha))$ by

$$1 - J_{h_n}(c_{\delta_n^r h_n}(1-\alpha)). \tag{4.1}$$

And it leads to the approximation of $\sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}(T_n(\theta_0) > c_{n,b}(1-\alpha))$ by

$$AsySz_n(\theta_0) = \sup_{h\in H}\left(1 - J_h(c_{\delta_n^r h}(1-\alpha))\right). \tag{4.2}$$

Suppose $J_h(c_g(1-\alpha))$ is a continuous function of (g,h) at each $(g,h) \in GH$ and Assumption C(ii) holds, i.e., $\delta_n = b/n \to 0$. Then, as $n \to \infty$ the quantity in (4.1) approaches $1 - J_h(c_0(1-\alpha))$ if $h_n \to h \in [0,\infty)$. It approaches $1 - J_\infty(c_g(1-\alpha))$ if $h_n \to \infty$ and $\delta_n^r h_n \to g \in [0,\infty]$. Hence, for any $(g,h) \in GH$, $\lim_{n\to\infty}(1-J_{h_n}(c_{\delta_n^r h_n}(1-\alpha))) = 1 - J_h(c_g(1-\alpha))$ for a suitable choice of $\{h_n \in H : n \geq 1\}$. This suggests that

$$\lim_{n\to\infty}\sup_{h\in H}(1 - J_h(c_{\delta_n^r h}(1-\alpha))) = \sup_{(g,h)\in GH}(1 - J_h(c_g(1-\alpha))) = AsySz(\theta_0). \tag{4.3}$$

It is shown below that (4.3) does hold, which implies that $AsySz_n(\theta_0)$ is an asymptotically valid finite-sample adjustment to $AsySz(\theta_0)$.

We now consider the general case in which γ may contain subvectors γ_2 and γ_3 and $p \geq 1$. In this case, only the subvector γ_1 affects whether γ is near a discontinuity point of the limit distribution. In consequence, only h_1, and not h_2, is affected by the δ_n^r rescaling that occurs above. For a subsampling test, we define

$$AsySz_n(\theta_0) = \sup_{h=(h_1,h_2)\in H}\left(1 - J_h(c_{(\delta_n^r h_1, h_2)}(1-\alpha))\right). \tag{4.4}$$

Next, we use the finite-sample adjustment to construct adjusted SC-Type and PSC-Type tests for Type = Sub and Hyb, which are denoted ASC-Type and APSC-Type tests. For $\delta \in (0,1)$ and $h_2 \in H_2$, define

$$\kappa(\delta,\alpha) = \sup_{h=(h_1,h_2)\in H}[c_{(h_1,h_2)}(1-\alpha) - c_{(\delta^r h_1,h_2)}(1-\alpha)],$$

$$\begin{aligned}
\kappa_{h_2}(\delta,\alpha) &= \sup_{h_1\in H_1}[c_{(h_1,h_2)}(1-\alpha)-c_{(\delta^r h_1,h_2)}(1-\alpha)],\\
\kappa^*(\delta,\alpha) &= \sup_{h\in H^*(\delta)} c_h(1-\alpha)-c_\infty(1-\alpha), \text{ and} \qquad (4.5)\\
\kappa^*_{h_2}(\delta,\alpha) &= \sup_{h_1\in H^*_{h_2}(\delta)} c_{(h_1,h_2)}(1-\alpha)-c_\infty(1-\alpha), \text{ where}\\
H^*(\delta) &= \{h\in H: c_{(\delta^r h_1,h_2)}(1-\alpha)<c_{(h_1,h_2)}(1-\alpha) \text{ for } h=(h_1,h_2)\},\\
H^*_{h_2}(\delta) &= \{h_1\in H_1: c_{(\delta^r h_1,h_2)}(1-\alpha)<c_{(h_1,h_2)}(1-\alpha)\}.
\end{aligned}$$

If $H^*(\delta)$ is empty, then $\kappa^*(\delta,\alpha)=-\infty$. If $H^*_{h_2}(\delta)$ is empty, then $\kappa^*_{h_2}(\delta,\alpha)=-\infty$. The ASC-Sub and ASC-Hyb tests are defined as in (3.1) with $\kappa(\alpha)$ and $\kappa^*(\alpha)$ replaced by $\kappa(\delta_n,\alpha)$ and $\kappa^*(\delta_n,\alpha)$, respectively, where $\delta_n = b/n$. The APSC-Sub and APSC-Hyb tests are defined as in (3.1) with $\kappa(\alpha)$ and $\kappa^*(\alpha)$ replaced by $\kappa_{\widehat{\gamma}_{2,n}}(\delta_n,\alpha)$ and $\kappa^*_{\widehat{\gamma}_{2,n}}(\delta_n,\alpha)$, respectively.

We use the following assumptions.

Assumption P. (i) The function $(g,h)\to J_h(c_g(1-\alpha))$ for $(g,h)\in H\times H$ is continuous at all $(g,h)\in GH$ and (ii) $Max_{Sub}(\alpha)=Max^-_{Sub}(\alpha)$, where the latter are defined as $Max_{Hyb}(\alpha)$ is defined in (2.14) but with $c_g(1-\alpha))$ in place of $\max\{c_g(1-\alpha),c_\infty(1-\alpha)\}$ and in addition $Max^-_{Sub}(\alpha)$ has $J_h(x)$ replaced by $J_h(x-)$.

Assumption Q. $c_h(1-\alpha)$ is continuous in h on H.

Assumption R. Either H^* is non-empty and $\sup_{h\in H^\dagger} c_h(1-\alpha)\le \sup_{h\in H^*} c_h(1-\alpha)$, where $H^\dagger=\{h\in H: h=\lim_{k\to\infty} h_{v_k}$ for some subsequence $\{v_k\}$ and some $h_{v_k}\in H^*(\delta_{v_k})$ for all $k\ge 1\}$, or H^* is empty and $H^*(\delta)$ is empty for all $\delta>0$ sufficiently close to zero.

Assumption S. For all $h_2\in H_2$, either $H^*_{h_2}$ is non-empty and $\sup_{h_1\in H^\dagger_{h_2}} c_{(h_1,h_2)}(1-\alpha)$ $\le \sup_{h_1\in H^*_{h_2}(\delta)} c_{(h_1,h_2)}(1-\alpha)$, where $H^\dagger_{h_2}=\{h_1\in H_1: h_1=\lim_{k\to\infty} h_{v_k,1}$ for some subsequence $\{v_k\}$ and some $h_{v_k,1}\in H^*_{\gamma_{v_k,2}}(\delta_{v_k})$ for all $k\ge 1$, where $\lim_{k\to\infty}\gamma_{v_k,2}=h_2\}$, or $H^*_{h_2}$ is empty and $H^*_{h_2}(\delta)$ is empty for all $\delta>0$ sufficiently close to zero.

Assumption P is a mild continuity assumption. Assumptions Q, R, and S are not restrictive in most examples. Whether Assumptions R and S hold depends primarily on the shape of $c_h(1-\alpha)$ as a function of h. It is possible for Assumptions R and S to be violated, but only for quite specific and unusual shapes for $c_h(1-\alpha)$. For example, Assumption R is violated in the case where $p=1$ and no parameter h_2 exists if for some $h^*\in(0,\infty)$ the graph of $c_h(1-\alpha)$ is (i) bowl-shaped for $h\in[0,h^*]$ with $c_0(1-\alpha)=c_{h^*}(1-\alpha)$ and (ii) strictly decreasing for $h>h^*$ with $c_\infty(1-\alpha)<c_h(1-\alpha)$ for all $0\le h<\infty$. In this case, H^* is empty (because $c_h(1-\alpha)$ takes on its minimum for $h=\infty$ and its maximum at $h=0$), but $h^*\in H^*(\delta)$ for all $\delta\in(0,1)$, which contradicts Assumption R.

The following result shows that $AsySz_n(\theta_0)$ provides an asymptotically valid finite-sample adjustment to $AsySz(\theta_0)$ that depends explicitly on the ratio $\delta_n = b/n$ and that the ASC and APSC tests have $AsySz(\theta_0)=\alpha$.

Theorem 4 (a) *Suppose Assumptions* A-G *and* P *hold. Then, a subsampling test satisfies*

$$\lim_{n\to\infty} AsySz_n(\theta_0) = AsySz(\theta_0).$$

(b) *Suppose Assumptions* A-G, K-M, Q, *and* R *hold. Then,* (i) $\lim_{n\to\infty}\kappa(\delta_n,\alpha) = \kappa(\alpha)$ *and* $\lim_{n\to\infty}\kappa^*(\delta_n,\alpha) = \kappa^*(\alpha)$ *and* (ii) *the* ASC-Sub *and* ASC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.
(c) *Suppose Assumptions* A-G, K, L, N, O, Q, *and* S *hold. Then,* (i) $\kappa_{\widehat{\gamma}_{n,2}}(\delta_n,\alpha) - \kappa_{\gamma_{n,2}}(\alpha) \to_p 0$ *and* $\kappa^*_{\widehat{\gamma}_{n,2}}(\delta_n,\alpha) - \kappa^*_{\gamma_{n,2}}(\alpha) \to_p 0$ *under all sequences* $\{\gamma_n = (\gamma_{n,1},\gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\}$ *and* (ii) *the* APSC-Sub *and* APSC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.

Comments. 1. An analogous result to Theorem 4(a) holds for the hybrid test with $c_{(\delta^r h_1,h_2)}(1-\alpha)$ replaced by $\max\{c_{(\delta^r h_1,h_2)}(1-\alpha), c_\infty(1-\alpha)\}$ in (4.4).

2. In Theorem 4(b), the ASC-Hyb test satisfies $\liminf_{n\to\infty}\kappa^*(\delta_n,\alpha) \geq \kappa^*(\alpha)$ and $AsySz(\theta_0) \leq \alpha$ without imposing Assumption R. Assumption R is a necessary and sufficient condition for $\lim_{n\to\infty}\kappa^*(\delta_n,\alpha) = \kappa^*(\alpha)$ given the other assumptions.

5 Equal-Tailed Tests

This section considers *equal-tailed* two-sided *hybrid* t tests. For brevity, equal-tailed SC, ..., APSC t tests are discussed in the Supplement to this paper. We suppose that $T_n(\theta_0) = \tau_n(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n$, where $\widehat{\theta}_n$ is an estimator of a scalar parameter θ based on a sample of size n, $\widehat{\sigma}_n$ $(\in R)$ is an estimator of the scale of $\widehat{\theta}_n$, and τ_n is a normalization constant, usually equal to $n^{1/2}$. An equal-tailed hybrid 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

$$\begin{aligned} T_n(\theta_0) &> c^*_{n,b}(1-\alpha/2) \text{ or } T_n(\theta_0) < c^{**}_{n,b}(\alpha/2), \text{ where} \\ c^*_{n,b}(1-\alpha/2) &= \max\{c_{n,b}(1-\alpha/2), c_\infty(1-\alpha/2)\} \text{ and} \\ c^{**}_{n,b}(\alpha/2) &= \min\{c_{n,b}(\alpha/2), c_\infty(\alpha/2)\}. \end{aligned} \tag{5.1}$$

Define

$$Max_{ET,Hyb}(\alpha) = \sup_{(g,h)\in GH} [1 - J_h(c^*_g(1-\alpha/2)) + J_h(c^{**}_g(\alpha/2))], \tag{5.2}$$

where $c^*_g(1-\alpha/2) = \max\{c_g(1-\alpha/2), c_\infty(1-\alpha/2)\}$ and $c^{**}_g(\alpha/2) = \min\{c_g(\alpha/2), c_\infty(\alpha/2)\}$.

If $J_h(x)$ is continuous at suitable (h,x) values, then the following assumption holds.

Assumption TET. $Max^{r-}_{ET,Hyb}(\alpha) = Max^{\ell-}_{ET,Hyb}(\alpha)$, where $Max^{r-}_{ET,Hyb}(\alpha)$ is defined as $Max_{ET,Hyb}(\alpha)$ is defined in (5.2) but with $J_h(c^{**}_g(\alpha/2)-)$ in place of

$J_h(c_g^{**}(\alpha/2))$ (where $J_h(c_g^{**}(\alpha/2)-)$ denotes the limit from the left of $J_h(x)$ at $x = c_g^{**}(\alpha/2)$) and $Max_{ET,Hyb}^{\ell-}(\alpha)$ is defined as in (5.2) with $J_h(c_g^*(1-\alpha/2)-)$ in place of $J_h(c_g^*(1-\alpha/2))$.

Assumption TET holds in the post-conservative model selection example by the continuity of $J_h^*(x)$ in x for $x \in R$ for all $h \in H$.

The proof of Theorem 1 of AG1 can be adjusted straightforwardly to yield the following result for equal-tailed hybrid t tests.

Corollary 1 *Let* $\alpha \in (0, 1/2)$ *be given. Let* $T_n(\theta_0) = \tau_n(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n$. *Suppose Assumptions* A-E, G, J, K, *and* TET *hold. Then, an equal-tailed hybrid* t *test satisfies* $AsySz(\theta_0) = Max_{ET,Hyb}(\alpha)$.

6 Confidence Intervals

This section introduces hybrid and size-corrected CIs for a parameter $\theta \in R^d$ when nuisance parameters $\eta \in R^s$ and $\gamma_3 \in \mathcal{T}_3$ may appear. (See Andrews and Guggenberger (2009d) for results concerning FCV and subsampling CIs.) The confidence level of a CI for θ requires uniformity over θ as well as over (η, γ_3). We make θ and η sub-vectors of γ so that the results from previous sections, which are uniform over $\gamma \in \Gamma$, give the uniformity results that we need for CIs for θ.[9]

Specifically, we partition θ into (θ_1, θ_2), where $\theta_j \in R^{d_j}$ for $j = 1, 2$, and we partition η into (η_1, η_2), where $\eta_j \in R^{s_j}$ for $j = 1, 2$. Then, we consider the same set-up as in Section 2.2 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 $q = d_2 + s_2$. 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 $s_2 = 0$).

We consider a test statistic $T_n(\theta_0)$ for testing the null hypothesis $H_0 : \theta = \theta_0$ as above. We obtain CIs for θ by inverting tests based on $T_n(\theta_0)$. Let Θ ($\subset R^d$) denote the parameter space for θ and let Γ denote the parameter space for γ. Hybrid CIs for θ are defined by

$$CI_n = \{\theta_0 \in \Theta : T_n(\theta_0) \leq c_{1-\alpha}\}, \tag{6.1}$$

where $c_{1-\alpha} = \max\{c_{n,b}(1-\alpha), c_\infty(1-\alpha)\}$. The critical value $c_{1-\alpha}$ does not depend on θ_0 when Assumption Sub1 holds, but does depend on θ_0 when Assumption Sub2 holds through the dependence of the subsample statistic on θ_0. 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 and Assumption Sub1 holds. Then, the corresponding CI of

[9] 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 sections on testing.

nominal level α is defined by

$$\begin{aligned} CI_n &= [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}, \infty),\ CI_n = (-\infty, \widehat{\theta}_n + \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}],\ \text{or} \\ CI_n &= [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}, \widehat{\theta}_n + \tau_n^{-1}\widehat{\sigma}_n c_{1-\alpha}], \end{aligned} \tag{6.2}$$

respectively (provided Θ is R).

The exact and asymptotic confidence sizes of CI_n are

$$ExCS_n = \inf_{\gamma\in\Gamma} P_\gamma(T_n(\theta) \le c_{1-\alpha}) \text{ and } AsyCS = \liminf_{n\to\infty} ExCS_n, \tag{6.3}$$

respectively, where $\theta = (\theta_1, \theta_2)$ and 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.

An equal-tailed hybrid CI for θ of nominal level α is defined by

$$CI_n = [\widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c^*_{n,b}(1-\alpha/2), \widehat{\theta}_n - \tau_n^{-1}\widehat{\sigma}_n c^{**}_{n,b}(\alpha/2)], \tag{6.4}$$

where $c^*_{n,b}(1-\alpha/2)$ and $c^{**}_{n,b}(\alpha/2)$ are defined in (5.1).

An analogue of Theorem 4 holds regarding the finite-sample-adjusted asymptotic sizes of subsampling and hybrid CIs. In this case, $AsyCS_n$ is defined as $AsySz_n$ is defined in (4.4) but with $\sup_{h\in H}$ replaced by $\inf_{h\in H}$ and J_h replaced by $1 - J_h$.

Next, we consider size-corrected CIs. SC-FCV, SC-Sub, and SC-Hyb CIs are defined by (6.1) with their critical values, $c_{1-\alpha}$, defined as in (3.1)-(3.2) for SC tests.

The following are changes in the assumptions for use with CIs.

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 $\{\gamma_{n,h} : n \ge 1\}$ being $\{P_{\theta_0,\gamma_{n,h}}(\cdot) : n \ge 1\}$, they are $\{P_{\gamma_{n,h}}(\cdot) : n \ge 1\}$. (iii) The test statistic $T_n(\theta_0)$ is replaced in the assumptions under a true sequence $\{\gamma_{n,h} : n \ge 1\}$ by $T_n(\theta_{n,h})$, where $\gamma_{n,h} = ((\theta_{n,h,1}, \eta_{n,h,1}), (\theta_{n,h,2}, \eta_{n,h,2}), \gamma_{n,h,3})$ 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 $\gamma = ((\theta_1, \eta_1), (\theta_2, \eta_2), \gamma_3)$. (v) θ_0 is replaced in the definition of $U_{n,b}(x)$ by θ_n when the true parameter is $\gamma_n = ((\theta_{n,1}, \eta_{n,1}), (\theta_{n,2}, \eta_{n,2}), \gamma_{n,3})$ and $\theta_n = ((\theta_{n,1}, \theta_{n,2})$.

Hybrid and size-corrected CIs satisfy the following results.

Corollary 2 *Let the assumptions be adjusted for* CI*s as stated above.*
(a) *Suppose Assumptions* A-G, K, *and* T *hold. Then, the hybrid* CI *satisfies* $AsyCS = 1 - Max_{Hyb}(\alpha)$.
(b) *Let* $\alpha \in (0, 1/2)$ *be given. Suppose Assumptions* A-E, G, J, K, *and* TET *hold. Then, the equal-tailed hybrid* t CI *satisfies* $AsyCS = 1 - Max_{ET,Hyb}(\alpha)$.
(c) *Suppose Assumptions* A-G *and* K-M *hold. Then, the* SC-FCV, SC-Sub, *and* SC-Hyb CI*s satisfy* $AsyCS = 1 - \alpha$.

Comment. Corollary 2(a), (b), and (c) hold by the same arguments as for Theorem 1, Corollary 1, and Theorem 2, respectively, with some adjustments.

Definitions and results for CIs of the form PSC-Type for Type = FCV, Sub, and Hyb, and ASC-Type and APSC-Type for Type = Sub and Hyb are analogous to those just stated for SC CIs but with critical values as defined in Sections 3.3 and 4, rather than as in Section 3.1. Size-corrected equal-tailed CIs are defined as in (6.4) with critical values $c_{1-\alpha/2}$ and $c_{\alpha/2}$ given by the equal-tailed SC, PSC, ASC, and/or APSC critical values for tests given in the Supplement in place of $c^*_{n,b}(1-\alpha/2)$ and $c^{**}_{n,b}(\alpha/2)$.

7 CI for an Autoregressive Parameter

We now apply the general results above to an AR(1) model with conditional heteroskedasticity. We use the unobserved components representations of the AR(1) model. The observed time series $\{Y_i : i = 0, ..., n\}$ is based on a latent no-intercept AR(1) time series $\{Y_i^* : i = 0, ..., n\}$:

$$\begin{aligned} Y_i &= \alpha + \beta i + Y_i^*, \\ Y_i^* &= \rho Y_{i-1}^* + U_i, \text{ for } i = 1, ..., n, \end{aligned} \tag{7.1}$$

where $\rho \in [-1+\varepsilon, 1]$ for some $0 < \varepsilon < 2$, $\{U_i : i = 0, \pm 1, \pm 2, ...\}$ are stationary and ergodic with conditional mean 0 given a σ-field $\mathcal{G}_{i-1}$ defined below, conditional variance $\sigma_i^2 = E(U_i^2|\mathcal{G}_{i-1})$, unconditional variance $\sigma_U^2 \in (0, \infty)$, and distribution F. The distribution of Y_0^* is the distribution that yields strict stationarity for $\{Y_i^* : i \leq n\}$ when $\rho < 1$, i.e., $Y_0^* = \sum_{j=0}^{\infty} \rho^j U_{-j}$, and is arbitrary when $\rho = 1$. We consider two versions of the AR(1) model—model 1, which has an intercept, and model 2, which has an intercept and time trend. Model 1 is obtained by setting $\beta = 0$ in (7.1). In the notation above, we have $\theta = 1 - \rho \in \Theta = [0, 2-\varepsilon]$.

Models (1) and (2) can be rewritten as

$$\begin{aligned} &(1)\ Y_i = \widetilde{\alpha} + \rho Y_{i-1} + U_i, \text{ where } \widetilde{\alpha} = \alpha(1-\rho), \text{ and} \\ &(2)\ Y_i = \overline{\alpha} + \overline{\beta} i + \rho Y_{i-1} + U_i, \text{ where } \overline{\alpha} = \alpha(1-\rho) + \rho\beta \text{ and } \overline{\beta} = \beta(1-\rho), \end{aligned} \tag{7.2}$$

for $i = 1, ..., n$.[10]

We consider a feasible quasi-GLS (FQGLS) t statistic based on estimators $\{\widehat{\phi}^2_{n,i} : i \leq n\}$ of the conditional variances $\{\sigma_i^2 : i \leq n\}$. The estimators $\{\widehat{\phi}^2_{n,i} : i \leq n\}$ may be based on a parametric specification of the conditional heteroskedasticity, such as a GARCH(1, 1) model, or a nonparametric procedure, such as one based on q lags of

[10] The advantage of writing the model as in (7.1) becomes clear here. For example, in model 1, the case $\rho = 1$ and $\widetilde{\alpha} \neq 0$ is automatically ruled out by model (7.1). This is a case where Y_i is dominated by a deterministic trend and the LS estimator of ρ converges at rate $n^{3/2}$.

the observations. In either case, we do not assume that the estimator of conditional heteroskedasticity is consistent. For example, we allow for incorrect specification of the parametric model in the former case and conditional heteroskedasticity that depends on more than q lags in the latter case. The estimated conditional variances $\{\widehat{\phi}_{n,i}^2 : i \leq n\}$ are defined such that they approximate a stationary $\mathcal{G}_{i-1}$-adapted sequence $\{\phi_i^2 : i \leq n\}$ in the sense that certain normalized sums have the same asymptotic distribution whether $\widehat{\phi}_{n,i}^2$ or ϕ_i^2 appears in the sum. This is a standard property of feasible and infeasible GLS estimators.

For example, for the model without a time trend, the results cover the case where (i) $\{\widehat{\phi}_{n,i}^2 : i \leq n\}$ are based on a GARCH(1,1) parametric model estimated using LS residuals with GARCH and LS parameter estimators $\widetilde{\pi}_n$ and $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$, respectively, (ii) $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$ have probability limit given by the true values $(\widetilde{\alpha}_0, \rho_0)$, see (7.2), (iii) $\widetilde{\pi}_n$ has a probability limit given by the "pseudo-true" value π_0, (iv) $\widehat{\phi}_{n,i}^2 = \phi_{i,1}^2(\widetilde{\alpha}_n, \widetilde{\rho}_n, \widetilde{\pi}_n)$, where $\phi_{i,1}^2(\widetilde{\alpha}, \rho, \pi)$ is the i-th GARCH conditional variance based on a start-up at time 1 and parameters $(\widetilde{\alpha}, \rho, \pi)$, and (v) $\phi_{i,-\infty}^2(\widetilde{\alpha}, \rho, \pi)$ is the GARCH conditional variance based on a start-up at time $-\infty$. In this case, $\phi_i^2 = \phi_{i,-\infty}^2(\widetilde{\alpha}_0, \rho_0, \pi_0)$. Thus, ϕ_i^2 is just $\widehat{\phi}_{n,i}^2$ with the estimation error and start-up truncation removed.

Under the null hypothesis that $\rho = \rho_n = 1 - \theta_n$, the studentized t statistic is

$$T_n^*(\theta_n) = \tau_n(\widehat{\rho} - \rho_n)/\widehat{\sigma}, \tag{7.3}$$

where $\tau_n = n^{1/2}$, $\widehat{\rho}$ is the LS estimator from the regression of $Y_i/\widehat{\phi}_i$ on $Y_{i-1}/\widehat{\phi}_i$ and $1/\widehat{\phi}_i$ in the case of model 1 and from the regression of $Y_i/\widehat{\phi}_i$ on $Y_{i-1}/\widehat{\phi}_i$, $1/\widehat{\phi}_i$, and $i/\widehat{\phi}_i$ in the case of model 2, and $\widehat{\sigma}^2$ is the $(1,1)$ element of the standard heteroskedasticity-robust variance estimator for the LS estimator in the preceding regression.

To define $T_n^*(\theta_n)$ more explicitly, let Y, U, X_1, and X_2 be n-vectors with ith elements given by $Y_i/\widehat{\phi}_i$, $U_i/\widehat{\phi}_i$, $Y_{i-1}/\widehat{\phi}_i$, and $1/\widehat{\phi}_i$, respectively, in models 1 and 2, except in model 2 let X_2 be the $n \times 2$ matrix with ith row $(1/\widehat{\phi}_i, i/\widehat{\phi}_i)$. Let Δ be the diagonal $n \times n$ matrix with ith diagonal element given by the ith element of the residual vector $M_X Y$, where $X = [X_1 : X_2]$ and $M_X = I_n - X(X'X)^{-1}X'$. That is, $\Delta = Diag(M_X Y)$. Then, by definition,

$$\begin{aligned} \widehat{\rho} &= \left(X_1' M_{X_2} X_1\right)^{-1} X_1' M_{X_2} Y, \text{ and} \\ \widehat{\sigma}^2 &= \left(n^{-1} X_1' M_{X_2} X_1\right)^{-1} \left(n^{-1} X_1' M_{X_2} \Delta^2 M_{X_2} X_1\right) \left(n^{-1} X_1' M_{X_2} X_1\right)^{-1}. \end{aligned} \tag{7.4}$$

For upper one-sided, lower one-sided, and symmetric two-sided tests or CIs concerning ρ, we take $T_n(\theta_n) = T_n^*(\theta_n)$, $-T_n^*(\theta_n)$, and $|T_n^*(\theta_n)|$, respectively.

In this section, we provide results for the (infeasible) QGLS estimator based on $\{\phi_i^2 : i \leq n\}$. Conditions under which feasible and infeasible QGLS estimators are asymptotically equivalent are technical and, for brevity, sufficient conditions are given in Andrews and Guggenberger (2008b). For technical reasons, these conditions take

$\widetilde{\pi}_n$ to be a discretized estimator and require $\widehat{\phi}_i^2$ to depend upon a finite number of lagged squared residuals. Neither of these conditions is particularly restrictive because the grid size for the discretized estimator can be defined such that there is little difference between the discretized and non-discretized versions of the estimator of π and any model with stationary conditional heteroskedasticity, such as a GARCH(1,1) model, can be approximated arbitrarily well by taking the number of lags sufficiently large.

By assumption, $\{(U_i, \phi_i^2) : i \geq 1\}$ are stationary and strong mixing. We define $\mathcal{G}_i$ to be some non-decreasing sequence of σ-fields for $i \geq 1$ for which $(U_j, \phi_{j+1}^2) \in \mathcal{G}_i$ for all $j \leq i$.

The vector of parameters is $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, where $\gamma_1 = \theta$ $(= 1 - \rho)$, $\gamma_2 = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7)' \in R^7$, where $\lambda_1 = Var_F(U_i)$, $\lambda_2 = Var_F(U_i/\phi_i^2) = E_F(\sigma_i^2/\phi_i^4)$, $\lambda_3 = Cov_F(U_i, U_i/\phi_i^2) = E_F(\sigma_i^2/\phi_i^2)$, $\lambda_4 = E_F\phi_i^{-1}$, $\lambda_5 = E_F\phi_i^{-2}$, $\lambda_6 = E_F\phi_i^{-4}$, and $\lambda_7 = Corr_F(U_i, U_i/\phi_i^2)) = \lambda_3/(\lambda_1\lambda_2)^{1/2}$, $\gamma_3 = (\alpha, F)$ in model 1, and $\gamma_3 = (\alpha, \beta, F)$ in model 2.[11] In this example, $\gamma_2 = \eta_2$ and no parameters θ_2 or η_1 appear. The distribution of the initial condition Y_0^* does not appear in γ_3 because under strict stationarity it equals the stationary marginal distribution of U_i and that is completely determined by F and γ_1 and in the unit root case it is irrelevant. In the definition of $\gamma_{n,h}$, we take $r = 1$.

The parameter spaces are $\Gamma_1 = \Theta = [0, 2 - \varepsilon]$ for some $\varepsilon > 0$, $\Gamma_2 \subset \Gamma_2^* = \{(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7) \in [\varepsilon_2, \infty)^2 \times (0, \infty) \times [\varepsilon_2, \infty)^3 \times [0, 1] : \lambda_7 = \lambda_3/(\lambda_1\lambda_2)^{1/2}\}$ for some $\varepsilon_2 > 0$, $\Gamma_3(\gamma_2) = B_1 \times \mathcal{F}(\gamma_2)$ in model 1, and $\Gamma_3(\gamma_2) = B_2 \times \mathcal{F}(\gamma_2)$ in model 2, where B_1 and B_2 are bounded subsets of R and R^2, respectively, and $\mathcal{F}(\gamma_2)$ is the parameter space for the stationary and strong-mixing distribution F of $\{U_i : i = ..., 1, 2, ...\}$ for a given value of γ_2.[12] In particular, we have

$$\begin{aligned}\mathcal{F}(\gamma_2) = \{F :\ & \{(U_i, \phi_i^2) : i = 0, \pm1, \pm2, ...\} \text{ are stationary and strong mixing} \\ & \text{under } F \text{ with } E_F(U_i|\mathcal{G}_{i-1}) = 0 \text{ a.s., } E_F(U_i^2|\mathcal{G}_{i-1}) = \sigma_i^2 \text{ a.s., where } \mathcal{G}_i \\ & \text{is some non-decreasing sequence of } \sigma\text{-fields for } i = ..., 1, 2, ... \text{ for} \\ & \text{which } (U_j, \phi_{j+1}^2) \in \mathcal{G}_i \text{ for all } j \leq i, \text{ the strong-mixing numbers} \\ & \{\alpha_F(m) : m \geq 1\} \text{ satisfy } \alpha_F(m) \leq Cm^{-3\zeta/(\zeta-3)} \text{ as } m \to \infty \text{ for some } \zeta > 3, \\ & \sup_{i,s,t,u,v,A} E_F|\textstyle\prod_{a\in A} a|^\zeta \leq M, \text{ where } 0 \leq i, s, t, u, v < \infty, \text{ and } A \text{ is any}\end{aligned}$$

[11] Note that Section 6 discusses CIs for θ, which is an element of γ, whereas here we consider CIs for $\rho = 1 - \theta$, which is not an element of γ. However, a CI for θ immediately yields one for ρ.

[12] The parameter space Γ_2 is a subset of Γ_2^* because the elements of γ_2 are related (given that they all depend on moments of (U_i, ϕ_i)) and Γ_2^* does not incorporate all of these restrictions. An example of a restriction is $\lambda_4^2 = (E_F\phi_i^{-1})^2 \leq E_F\phi_i^{-2} \cdot E_F 1 = \lambda_5$ by the Cauchy-Schwarz inequality. Although the restrictions on Γ_2 are not written explicitly, this is not a problem because the subsampling and hybrid procedures do not depend on the specification of Γ_2 and the size-correction procedures only depend on λ_7 or $h_{2,7}$ whose parameter space is known.

The parameter space B_1 is taken to be bounded, because otherwise there are sequences $\alpha_n \to \infty$, $\rho_n \to 1$ for which $\widetilde{\alpha}_n \nrightarrow 0$. For analogous reasons, B_2 is taken to be bounded.

$$
\begin{aligned}
&\text{nonempty subset of } \{U_{i-s}, U_{i-t}, U_{i+1}^2, U_{-u}, U_{-v}, U_1^2\},\ \phi_i^2 \geq \delta \ a.s.,\\
&\lambda_{\min}(E_F X^1 X^{1\prime} U_1^2/\phi_1^2) \geq \delta,\ \text{where } X^1 = (Y_0^*/\phi_1, \phi_1^{-1})',\\
&Var_F(U_i) = \lambda_1, Var_F(U_i/\phi_i^2) = \lambda_2, Cov_F(U_i, U_i/\phi_i^2) = \lambda_3,\\
&E_F\phi_i^{-1} = \lambda_4, E_F\phi_i^{-2} = \lambda_5, E_F\phi_i^{-4} = \lambda_6,\ \text{and } Corr_F(U_i, U_i/\phi_i^2))'\\
&= \lambda_7,\ \text{where } \gamma_2 = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7)'\} \qquad (7.5)
\end{aligned}
$$

for some $C, M < \infty$ and $\delta > 0$, where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of a matrix A.

In the Supplement, we verify the assumptions of Corollary 2 concerning hybrid CIs, except Assumption B. The Supplement uses Lemma 4 of AG1 to verify Assumption G. Assumption B holds by Theorem 1 in Andrews and Guggenberger (2008b). The slightly weaker assumptions than those in Corollary 2 yield asymptotic size results for FCV and subsampling CIs, see Theorem 3 in Andrews and Guggenberger (2009d). For brevity, we verify assumptions only for model 1. The moment conditions in $\mathcal{F}(\gamma_2)$ are used in the verification of Assumptions B and E for the case where $\rho \to 1$ at a rate slower than n^{-1}. The bounding of ϕ_i^2 away from zero in $\mathcal{F}(\gamma_2)$ is not restrictive because it is a consequence of a suitable choice of $\widehat{\phi}_i^2$.

In this example, $H = R_{+,\infty} \times \Gamma_2$. Therefore, to establish Assumption B, we have to consider sequences $\{\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3})' : n \geq 1\}$, where $h = (h_1, h_2)$, when the true autoregressive parameter $\rho = \rho_n$ equals $1 - \gamma_{n,h,1}$ where (i) $h_1 = \infty$ and (ii) $0 \leq h_1 < \infty$. For AR(1) models with conditional heteroskedasticity, the special case of case (ii) in which $\rho = 1$ is fixed has been considered by Seo (1999) and Guo and Phillips (2001). For models without conditional heteroskedasticity, case (i) is studied by Park (2002), Giraitis and Phillips (2006), and Phillips and Magdalinos (2007), and case (ii) is the "near integrated" case that has been studied without conditional heteroskedasticity by Bobkowski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987), Elliott (1999), Elliott and Stock (2001), and Müller and Elliott (2003). The latter three papers consider the situation of interest here in which the initial condition Y_0^* yields a stationary process. Specifically, what is relevant here is the triangular array case with row-wise strictly stationary observations $\{Y_i^* : i \leq n\}$ and ρ that depends on n. Note that case (ii) contains as a special case the unit root model $\rho = 1$. We do not consider an AR(1) model here without an intercept, but such a model can be analyzed using the results of Andrews and Guggenberger (2008a). Interestingly, the asymptotic distributions in this case are quite different than in the models with an intercept or intercept and time trend.

For model 1, we have

$$
\begin{aligned}
&T_n^*(\theta_n) \to_d \ J_h^* \text{ under } \gamma_{n,h},\ \text{where}\\
&J_h^* \text{ is the } N(0,1) \text{ distribution for } h_1 = \infty,\\
&J_h^* \text{ is the distribution of } \left(h_{2,7} \frac{\int_0^1 I_{D,h}^*(r) dW(r)}{(\int_0^1 I_{D,h}^*(r)^2 dr)^{1/2}} + (1 - h_{2,7}^2)^{1/2} Z_2 \right)
\end{aligned}
$$

$$
\begin{aligned}
&\text{for } 0 \le h_1 < \infty, \\
&I^*_{D,h}(r) = I^*_h(r) - \int_0^1 I^*_h(s)ds, \\
&I^*_h(r) = I_h(r) + \frac{1}{\sqrt{2h_1}}\exp(-h_1 r)Z_1 \text{ for } h_1 > 0 \text{ and } I^*_h(r) = W(r) \text{ for } h_1 = 0, \\
&I_h(r) = \int_0^r \exp(-(r-s)h_1)dW(s),
\end{aligned} \tag{7.6}
$$

$W(\cdot)$ is a standard Brownian motion, and Z_1 and Z_2 are independent standard normal random variables that are independent of $W(\cdot)$. As defined, $I_h(r)$ is an Ornstein-Uhlenbeck process. The parameter $h_{27} \in [0,1]$ is the limit of $Corr_{F_n}(U_i, U_i/\phi_i^2))$ under the sequence $\{\gamma_{n,h} : n \ge 1\}$.

For model 2, (7.6) holds except that for $0 \le h_1 < \infty$ J^*_h is the distribution of

$$
h_{2,7}\frac{\int_0^1 \left[I^*_{D,h}(r) - 12\int_0^1 I^*_{D,h}(s)sds \cdot (r - 1/2)\right] dW(r)}{\left(\int_0^1 \left[I^*_{D,h}(r) - 12\int_0^1 I^*_{D,h}(s)sds \cdot (r-1/2)\right]^2 dr\right)^{1/2}} + (1 - h^2_{2,7})^{1/2}Z_2. \tag{7.7}
$$

The asymptotic results above apply to a first-order AR model. They should extend without essential change to CIs for the "sum of the AR coefficients" in a p-th order autoregressive model. In particular, the asymptotic distributions for statistics concerning the "sum of the AR coefficients" should be the same as those for ρ given in (7.6) and (7.7). Of course, the proofs will be more complex. For brevity, we do not provide such proofs.

Figure 1 provides .95 quantile graphs of J^*_h, $-J^*_h$, and $|J^*_h|$ as functions of h_1 for the cases of $h_{27} = 0$, .3, .6, and 1. The graphs for different values of h_{27} have similar shapes, but are progressively less steep as h_{27} decreases from 1 to 0. All of the graphs are monotone in h_1.[13] The .95 quantile graphs for J^*_h are monotone increasing in h_1 for each value of h_2 because the upper tail of J^*_h gets thinner as h_1 gets smaller. In consequence, the upper one-sided and equal-tailed two-sided subsampling CIs under-cover the true value asymptotically and the upper FCV CI has correct size asymptotically. The .95 quantile graphs for $-J^*_h$ are decreasing in h_1 for each value of h_2 because the lower tail of J^*_h gets thicker as h_1 gets smaller. The .95 quantile graphs for $|J^*_h|$ are decreasing in h_1 for each value of h_2 because the lower tail of J^*_h gets thicker as h_1 gets smaller at a faster rate than the upper tail of J^*_h gets thinner. Because the graphs of $-J^*_h$ and $|J^*_h|$ are decreasing in h_1, the lower and symmetric subsampling CIs have correct asymptotic size, while the lower FCV CI under-covers the true value asymptotically. These results explain the seemingly puzzling result (quantified in Table II below) that the equal-tailed subsampling CI has incorrect size asymptotically while the symmetric subsampling CI has correct size asymptotically.

[13] The graphs in Figures 1 and 2 are computed by simulation. Monotonicity in Figure 1 is established numerically, not analytically.

Table II reports the asymptotic (Asy) and finite-sample-adjusted asymptotic (Adj-Asy) sizes of nominal 95% CIs for model 1 for symmetric and equal-tailed two-sided FCV, subsampling, and hybrid CIs, see the last two rows of the Table. (Symmetric and equal-tailed FCV CIs are the same, so only the former are reported.) These numbers are obtained by simulating the asymptotic formulae of Section 6. Further details concerning the construction of Table II are given in the Supplement.

Table II also reports finite-sample coverage probabilities of these CIs based on a FQGLS estimator $\widehat{\rho}_n$ that uses a GARCH(1, 1) specification for the conditional heteroskedasticity. The GARCH parameters are estimated by the closed-form estimator of Kristensen and Linton (2006). This estimator is employed in the Monte Carlo simulations because it is very quick to compute. Six different forms of the true conditional heteroskedasticity of the innovations are considered: (i) GARCH(1, 1) with (intercept, MA, AR) parameters equal to (.20, .15, .80), (ii) IGARCH(1, 1) with (intercept, MA, AR) parameters (.20, .20, .80), (iii) GARCH(1, 1) with (intercept, MA, AR) parameters (.20, .70, .20), (iv) i.i.d., (v) ARCH(4) with (intercept, AR1-AR4) parameters (.20, .30, .20, .20, .20), and (vi) IARCH(4) with (intercept, AR1-AR4) parameters (.20, .30, .30, .20, .20). In all cases, $U_i = \sigma_i \varepsilon_i$, where ε_i is standard normal and σ_i is the multiplicative conditional heteroskedasticity. The ARCH and IARCH processes provide evidence concerning the robustness of the procedures to an incorrect specification of the form of the conditional heteroskedasticity used in the definition of $\widehat{\rho}_n$. The integrated GARCH and ARCH processes are not covered by the asymptotic results but are included to address questions of robustness. The sample size, subsample size, and number of subsamples of consecutive observations employed are $n = 131$, 12, and 119. (We did not experiment with other sample sizes or subsample sizes.)

For case (i), we report the finite-sample coverage probability for eight values of ρ between $-.9$ and 1.0, as well as the minimum over $\rho \in [-.9, 1]$, which is denoted FS-Min.[14] For brevity, for cases (ii)-(vi), we only report FS-Min. The finite-sample size of a CI depends on the minimum coverage probability over both ρ and different true forms of conditional heteroskedasticity. We do not attempt to determine the finite-sample size via simulations. For the four non-integrated cases, we report the asymptotic and finite-sample-adjusted asymptotic sizes that correspond to the particular value of h_{27} for the given case (which are $h_{27} = .86, .54, 1.0$, and .54 for cases (i), (iii), (iv), and (v), respectively). These are the correct asymptotic sizes if h_{27} is known. Table B-II in the Supplement reports results analogous to those in Table II for upper and lower one-sided CIs. We do not report results for any other CIs in the literature because none have correct asymptotic size.

We now discuss the results in Table II. The two-sided FCV CI under-covers asymptotically by a substantial amount, see the rows labeled Asy of column 4. Its $AsyCS$ equals 69.5%. The asymptotic results for equal-tailed subsampling CIs are similar,

[14] The minimum is calculated over the set $\{-.9, -.8, ..., .9, .95, .97, .99, 1.0\}$. The reason for excluding $\rho \in (-1.0, -.9)$ is discussed in the Supplement.

but somewhat worse, see the rows labeled Asy of column 7. Its $AsyCS$ equals 59.8%. Hence, subsampling CIs can have very poor asymptotic performance. On the other hand, symmetric subsampling CIs have correct $AsyCS$ (up to simulation error) for the reasons described above, see the second last row of column 5.

The discussion above of the quantile graphs in Figure 1 leads to the following results, which are corroborated by the numerical results. The two-sided FCV CI under-covers because its upper endpoint is farther away from 1 than it should be. Hence, it misses the true value of ρ too often to the left. On the other hand, the equal-tailed subsampling CI under-covers ρ because its lower endpoint is closer to 1 than it should be. Hence, it misses the true ρ to the right too often.

The finite-sample adjusted asymptotic results for $\delta_n = b/n = 12/131$ show much less severe under-rejection for the equal-tailed subsampling CIs than the unadjusted asymptotic results, compare the rows denoted Asy and Adj-Asy in column 7 of Table II. The finite-sample coverage probabilities of the subsampling CIs (for $n = 131$ and $b = 12$) are closer in most cases to the adjusted asymptotic sizes than the unadjusted asymptotic sizes. Hence, it is apparent that the asymptotic size of the equal-tailed subsampling CIs is approached slowly as $n \to \infty$ and is obtained only with large sample sizes. In consequence, increases in the sample size from $n = 131$ makes the equal-tailed subsampling CIs perform worse rather than better. The equal-tailed subsampling CIs can be size-corrected. We do not report results here for such CIs.

None of the CIs considered here are similar asymptotically in a uniform sense or in finite samples. For the latter, see the rows corresponding to different values of ρ in case (i).

The symmetric and equal-tailed hybrid CIs both have correct $AsyCS$, see the rows labeled Asy in columns 6 and 8 of Table II. This occurs because for every value of h_{27} either the critical value of the FCV CI or the subsampling CI is suitable. Hence, the maximum of the two is a critical value that delivers correct asymptotic size. The finite-sample minimum (over ρ) coverage probabilities of the symmetric hybrid CI over the six cases range from 94.6 to 96.2%, which is quite good given the wide variety of conditional heteroskedasticity covered by these six cases. For the equal-tailed hybrid CI, the range is 93.5 to 93.9%, which is also good, but slightly lower than desirable. It is far superior to that of the FCV or equal-tailed subsampling CIs.

We conclude by noting that the same sort of size issues that arise with subsampling in the AR(1) model also arise in vector autoregressive models with roots that may be near unity. For example, they arise with subsampling tests of Granger causality in such models, see Choi (2005).

8 Post-Conservative Model Selection Inference

Figure 2 provides graphs of the quantiles, $c_h(1-\alpha)$, of $|J_h^*|$ as a function of $h_1 \geq 0$ for several values of $h_2 \geq 0$ for this example. (The quantile graphs are invariant to the signs of h_1 and h_2.) The corresponding quantile graphs for J_h^* are remarkably similar to those for $|J_h^*|$ and, hence, are not given. In Figure 2, the graphs are hump shaped with the size of the hump increasing in $|h_2|$. Based on the shape of the graphs, one expects the subsampling, FCV, and hybrid tests all to over-reject the null hypothesis asymptotically and in finite samples and by a substantial amount when $|h_2|$ is large.

Table III provides null rejection probability results that are analogous to those in Table II but for the present example. The parameter space H_2 for the (asymptotic) correlation h_2 between the LS estimators of the two regressors is $[-.995, .995]$. The finite-sample results in Table III are for $n = 120$, $b = 12$, and a model with standard normal errors, and $k = 3$ regressors, where $x_{1,i}^*$ and $x_{2,i}^*$ are independent standard normal random variables and $x_{3,i}^* = 1$. To dramatically increase computational speed, finite-sample results for tests that utilize subsampling critical values are based on $q_n = 119$ subsamples of consecutive observations. Hence, only a small fraction of the "120 choose 12" available subsamples are used. In cases where such tests have correct asymptotic size, their finite-sample performance is expected to be better when all available subsamples are used than when only $q_n = 119$ are used. Further details concerning Table III are given in the Supplement.

The asymptotic results for the Sub, FCV, and Hyb tests show that all of these tests perform very similarly and very poorly. They are found to over-reject the null hypothesis very substantially for the upper and symmetric cases when the absolute value of the correlation, $|h_2|$, is large. (Results for equal-tailed tests, not reported, are similar to those for symmetric tests.) The asymptotic sizes of these nominal 5% tests range from 93 to 96% (see columns 2, 4, and 7). Even for $|h_2| = .8$, the maximum (over h_1) asymptotic rejection probabilities of these tests range from 36 to 44%. Adjusted asymptotic sizes of the nominal 5% Sub and Hyb tests, not reported, are slightly lower than the unadjusted ones, but they are still in the range of 90 to 92%.

The finite-sample maximum (over h_1 and h_2) null rejection probabilities of the nominal 5% Sub, FCV, and Hyb tests are very high and reflect the asymptotic results (see columns 3, 5, and 8).[15] They range from 91 to 95%.

Next, we consider PSC tests. We use the following consistent estimator of $\gamma_{n,2}$:

$$\widehat{\gamma}_{n,2} = \frac{-n^{-1}\sum_{i=1}^n x_{1i}x_{2i}}{(n^{-1}\sum_{i=1}^n x_{1i}x_{1i}n^{-1}\sum_{i=1}^n x_{2i}x_{2i})^{1/2}}, \tag{8.1}$$

where $\{(x_{1i}, x_{2i}) : i = 1, ..., n\}$ are the residuals from the regressions of x_{ji}^* on x_{3i}^* for $j = 1, 2$. The choice of this estimator is based on the equality $\gamma_{n,2} = Q_n^{12}/(Q_n^{11}Q_n^{22})^{1/2} = -Q_{n,12}/(Q_{n,11}Q_{n,22})^{1/2}$, where Q_n^{jm} and $Q_{n,jm}$ denote the (j, m) elements of Q_n^{-1} and $Q_n = E_{G_n}x_i^{\perp}x_i^{\perp\prime}$, respectively, for $j, m = 1, 2$ (see the second equality in (11.19)

[15] Strictly speaking, h_2 denotes the asymptotic correlation between the LS estimators and H_2 denotes its parameter space. For simplicity, when discussing the finite-sample results, we let h_2 denote the finite-sample correlation between the LS estimators and we let H_2 denote its parameter space.

of the Supplement) and G_n is the distribution of (ε_i, x_i^*). Consistency of $\widehat{\gamma}_{n,2}$ (i.e., $\widehat{\gamma}_{n,2} - \gamma_{n,2} \rightarrow_p 0$ under $\{\gamma_n : n \geq 1\}$) follows from a Lemma in the Supplement to this paper. Thus, Assumption N holds. Note that the PSC tests do not depend on the specification of the parameter space for h_2. The PSC-FCV CI obtained by inverting the PSC-FCV test considered here is closely related to, but different from, the modified CI of Kabaila (1998).

Table III reports finite-sample maximum (over h_1) null rejection probabilities of the PSC-FCV and PSC-Hyb tests (see columns 6 and 9). These tests both perform very well. The maximum (over h_1 and $|h_2|$) null rejection probabilities of these tests are all in the range of 4.8 to 5.3% for upper and symmetric tests. For both tests, the maximum rejection rates (over h_1) do not vary too much with $|h_2|$, which is the objective of the "plug-in" approach. Hence, the "plug-in" approach works well in this example.

For $H_2 = [-.999, .999]$, the finite-sample maximum (over h_1 and $|h_2|$) null rejection rates of the PSC tests lie between 6.9 and 7.4% For $H_2 = [-.9999, .9999]$, the PSC tests have corresponding values between 71 and 83%. Hence, it is clear that bounding $|h_2|$ away from 1.0 is not only sufficient for the asymptotic PSC results to hold, but it is necessary for the PSC tests to have good finite-sample size. For practical purposes, this is not much of a problem because (i) h_2 can be consistently estimated, so one has a good idea of whether $|h_2|$ is close to 1.0 and (ii) $|h_2|$ can be very close to 1.0 (i.e., .995 or less) and the PSC tests still perform very well in finite samples.

In conclusion, nominal 5% subsampling, FCV, and hybrid tests have asymptotic and adjusted-asymptotic sizes that are very large—between 90 and 96%—for upper, symmetric, and equal-tailed tests (for $H_2 = [-.995, .995]$). The maximum (over the cases considered) finite-sample null rejection probabilities of these tests for $n = 120$ and $b = 12$ are close to the asymptotic values. PSC methods work very well in this example. The PSC-Hyb and PSC-FCV tests have finite-sample maximum (over the cases considered) null rejection probabilities between 4.8 and 5.3% for upper, lower, and symmetric tests (for H_2 as above).

References

Anderson, T. W. and H. Rubin (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *Annals of Mathematical Statistics,* 21, 570-582.

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

Andrews, D. W. K. and H.-Y. Chen (1994): "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics*, 12, 187–204.

Andrews, D. W. K. and P. Guggenberger (2005): "Applications of Subsampling, Hybrid, and Size-Correction Methods," Cowles Foundation Discussion Paper No. 1608, Yale University.

——— (2008a): "Asymptotics for Stationary Very Nearly Unit Root Processes," *Journal of Time Series Analysis,* 29, 203-212.

——— (2008b): "Asymptotics for LS, GLS, and Feasible GLS Statistics in an AR(1) Model with Conditional Heteroskedaticity," Cowles Foundation Discussion Paper No. 1665, Yale University.

——— (2009a): "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap," *Econometric Theory,* 25, forthcoming.

——— (2009b): "Incorrect Asymptotic Size of Subsampling Procedures Based on Post-Consistent Model Selection Estimators," provisionally accepted at *Journal of Econometrics.*

——— (2009c): "Supplement to 'Hybrid and Size-Corrected Subsampling Methods,"' available on the Econometric Society website.

——— (2009d): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory,* 25, forthcoming.

Bobkowski, M. J. (1983): "Hypothesis Testing in Nonstationary Time Series," unpublished Ph. D. Thesis, University of Wisconsin, Madison.

Cavanagh, C. (1985): "Roots Local to Unity," unpublished manuscript, Harvard University.

Chan, N. H. and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics,* 15, 1050-1063.

Chen, W. and R. Deo (2007): "A Smooth Transition to the Unit Root Distribution Via the Chi-square Distribution with Interval Estimation for Nearly Integrated Autoregressive Processes," unpublished manuscript, Stern School of Business, New York University.

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.

Elliott, G. (1999): "Efficient Tests for a Unit Root When the Initial Observation is Drawn from Its Unconditional Distribution," *International Economic Review,* 40, 767-783.

Elliott, G. and J. H. Stock (2001): "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics,* 103, 155-181.

Giraitis, L. and P. C. B. Phillips (2006): "Uniform Limit Theory for Stationary Autoregression," *Journal of Time Series Analysis,* 27, 51-60. Corrigendum, forthcoming.

Guggenberger, P. (2009): "The Impact of a Hausman Pretest on the Asymptotic Size of a Hypothesis Test," *Econometric Theory,* 25, forthcoming.

Guggenberger, P. and R. J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak, and Strong Identification," *Econometric Theory,* 21, 667-709.

——— (2008): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," *Journal of Econometrics,* 142, 134-161.

Guo, B. and P. C. B. Phillips (2001): "Testing for Autocorrelation and Unit Roots in the Presence of Conditional Heteroskedasticity of Unknown Form," unpublished working paper, Department of Economics, UC Santa Cruz.

Hansen, B. E. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economic Studies,* 81, 594-607.

Jansson, M. and M. J. Moreira (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica,* 74, 681-714.

Kabaila, P. (1998): "Valid Confidence Intervals in Regression after Variable Selection," *Econometric Theory,* 14, 463-482.

Kim, K. and P. Schmidt (1993): "Unit Root Tests with Conditional Heteroskedasticity," *Journal of Econometrics,* 59, 287-300.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica,* 70, 1781-1803.

——— (2005): "Testing Parameters in GMM without Assuming They Are Identified," *Econometrica,* 73, 1103-1124.

Kristensen, D. and O. Linton (2006): "A Closed-Form Estimator for the GARCH(1,1) Model," *Econometric Theory,* 22, 323-337.

Leeb, H. (2006): "The Distribution of a Linear Predictor After Model Selection: Unconditional Finite-Sample Distributions and Asymptotic Approximations," *2nd Lehmann Symposium—Optimality.* Institute of Mathematical Statistics Lecture Notes—Monograph Series.

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

Mikusheva, A. (2007a): "Uniform Inferences in Autoregressive Models," *Econometrica,* 75, 1411-1452.

——— (2007b): "Supplement to 'Uniform Inferences in Autoregressive Models'," *Econometrica,* 75, 1411-1452. Available on the Econometric Society website.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica,* 71, 1027-1048.

——— (2009): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," *Journal of Econometrics,* forthcoming.

Müller, U. K. and G. Elliott (2003): "Tests for Unit Roots and the Initial Condition," *Econometrica,* 71, 1269-1286.

Nankervis, J. C. and N. E. Savin (1996): "The Level and Power of the Bootstrap t-test in the Trend Model with AR(1) Errors," *Journal of Business and Economic Statistics,* 14, 161-168.

Otsu, T. (2006): "Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models under Weak Identification," *Econometric Theory,* 22, 513-527.

Park, J. Y. (2002): "Weak Unit Roots," unpublished working paper, Department of Economics, Rice University.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika,* 74, 535-547.

Phillips, P. C. B. and T. Magdalinos (2007): "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics,* 136, 115-130.

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.

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

Seo, B. (1999):"Distribution Theory for Unit Root Tests with Conditional Heteroskedasticity," *Journal of Econometrics,* 91, 113-144.

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.

TABLE I

Asymptotic Sizes of Subsampling and Hybrid Tests and Confidence Intervals of Symmetric and Equal-Tailed Two-Sided Types for a Variety of Models*

(a) Nominal 5% Tests				
	Subsampling		Hybrid	
Model	Sym	Eq-Tail	Sym	Eq-Tail
Nuisance Parameter Near Boundary	10	52.5	5	5
Post-Conservative Model Selection	94	94	94	94
IV Regression–2SLS w/ Possibly Weak IVs	5.5	82	5	5
Parameter-Dependent Support	5	5	5	5
(b) Nominal 95% Confidence Intervals				
AR w/ Intercept	95	60	95	95
AR w/Intercept & Trend	95	25	95	95
Post-Consistent Model Selection	0	0	0	0
Parameter of Interest Near Boundary	90	47.5	95	95
Parameters Defined by Moment Inequalities	95	95	95	95

*The details of the "Post-Conservative Model Selection" model are described above in the text. The results for the "IV Regression–2SLS w/ Possibly Weak IVs" model are for tests concerning the coefficient on an endogenous variable in a linear IV model with a single endogenous variable and five IVs.

TABLE II

AR EXAMPLE: CI COVERAGE PROBABILITIES (×100) FOR NOMINAL 95% CIs*

Case	DGP	n=131 or *Asy*	Symmetric CIs FCV	Symmetric CIs Sub	Symmetric CIs Hyb	Equal-tailed CIs Sub	Equal-tailed CIs Hyb
		-.90	90.7	95.5	95.8	93.0	94.7
		-.50	91.0	95.0	95.7	90.7	93.6
		.00	91.1	96.0	96.4	90.2	94.1
(i)	GARCH	ρ =.70	90.4	97.9	97.9	88.7	96.6
	MA=.15,	.80	89.6	97.8	97.8	88.7	97.0
	AR=.80	.90	87.9	97.9	97.9	89.5	97.5
	h_{27} = .86	.97	82.1	97.5	97.5	92.7	98.0
		1.0	65.1	95.1	95.1	94.5	96.7
		FS-Min	65.1	95.0	95.1	88.4	93.6
		Asy	76.8	95.0	95.0	69.6	95.0
		Adj-Asy	-	95.0	95.1	89.0	95.1
(ii)	IGARCH						
	MA=.20,	FS-Min	67.3	95.4	95.6	87.8	93.5
	AR=.80						
(iii)	GARCH	FS-Min	70.4	95.4	96.0	89.6	93.9
	MA=.70,	Asy	87.5	95.0	95.0	85.2	95.0
	AR=.20	Adj-Asy	-	94.7	95.1	92.6	95.2
	h_{27} = .54						
(iv)	i.i.d.	FS-Min	62.4	94.6	94.6	88.1	93.5
	$h_{27} = 1$	Asy	69.4	95.0	95.0	59.7	95.0
		Adj-Asy	-	95.0	95.0	86.3	95.1
(v)	ARCH4	FS-Min	69.3	95.8	96.1	88.7	93.8
	(.3,.2,.2,.2)	Asy	87.5	95.0	95.0	85.2	95.0
	h_{27} = .54	Adj-Asy	-	94.7	95.1	92.6	95.2
(vi)	IARCH4	FS-Min	71.1	95.7	96.2	88.5	93.7
	(.3,.3,.2,.2)						
	Min Over	Asy	69.5	94.9	94.9	59.8	95.0
	$h_{27} \in [0, 1]$	Adj-Asy	-	94.6	95.0	86.3	95.1

*The asymptotic and adjusted asymptotic results of Table II are based on 30,000 simulation repetitions. The search over h_1 to determine the minimum is done on the interval $[-.90, 1]$ with stepsize 0.01 on $[-.90, .90]$ and stepsize .001 on $[.90, 1.0]$. The search over h_{27} to determine the minimum is done on the interval $[0, 1]$ with stepsize 0.05. The asymptotic results are computed using a discrete approximation to the continuous stochastic process on $[0, 1]$ with 25,000 grid points.

TABLE III

CONSERVATIVE MODEL SELECTION EXAMPLE: MAXIMUM (OVER h_1) NULL REJECTION PROBABILITIES ($\times$100) FOR DIFFERENT VALUES OF THE CORRELATION h_2 FOR VARIOUS NOMINAL 5% TESTS, WHERE THE PROBABILITIES ARE ASYMPTOTIC, FINITE-SAMPLE-ADJUSTED ASYMPTOTIC, AND FINITE SAMPLE FOR $n = 120$ AND $b = 12$ AND THE PARAMETER SPACE FOR h_2 IS $[-.995, .995]^*$

(a) Upper 1-Sided Tests

	Test:	Sub	Sub	FCV	FCV	PSC-FCV	Hyb	Hyb	PSC-Hyb
$\lvert h_2\rvert$	Prob:	Asy	n=120	Asy	n=120	n=120	Asy	n=120	n=120
.00		5.1	5.4	5.3	5.4	4.7	5.1	3.7	3.3
.20		6.9	7.2	7.1	7.5	5.1	6.9	5.3	4.0
.40		11.2	11.0	11.8	11.9	5.1	11.2	8.7	4.5
.60		20.2	19.8	21.8	22.0	4.9	20.2	17.3	4.8
.80		41.3	38.9	44.3	43.8	4.8	41.3	37.2	4.8
.90		61.3	57.5	63.9	62.8	4.6	61.3	56.8	4.6
.95		75.5	72.2	77.2	76.7	4.6	75.5	71.8	4.6
.995		92.9	91.9	93.2	93.1	4.1	92.9	91.9	4.1
Max		92.9	91.9	93.2	93.1	5.1	92.9	91.9	4.8

(b) Symmetric 2-Sided Tests

.00		5.1	5.0	5.4	5.5	5.0	5.1	3.3	3.1
.20		6.0	5.3	6.3	6.5	5.1	6.0	3.8	3.3
.40		8.7	7.3	9.6	10.1	5.2	8.7	5.9	4.0
.60		16.1	12.3	18.2	18.8	5.3	16.1	11.3	4.8
.80		36.2	28.2	40.6	40.3	4.9	36.2	27.8	4.8
.90		57.6	48.5	62.0	61.5	4.5	57.6	48.3	4.5
.95		73.4	66.1	77.1	76.4	4.2	73.4	66.0	4.2
.995		93.9	90.7	95.5	95.3	4.2	93.9	90.7	4.2
Max		93.9	90.7	95.5	95.3	5.3	93.9	90.7	4.8

*The results in Table III are based on 20,000 simulation repetitions. For the finite-sample results, the search over $|\beta_2|$ is done on the interval $[0, 10]$ with stepsizes 0.0025, 0.025, and .250, respectively, on the intervals $[0.0, 0.8]$, $[0.8, 3]$, and $[3, 10]$ and also includes the value $|\beta_2| = 999,999$. For the asymptotic results the search over $|h_1|$ is done in the interval $[-10, 10]$ with stepsize 0.01. For the finite-sample and asymptotic results, the Max is taken over $|h_2|$ values in $\{0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 0.99, 0.995\}$. For the plug-in size-correction values, the grid of $|\gamma_2|$ values has stepsizes .01, .001, .0001, and .00002, respectively, on the intervals $[0.0, 0.7]$, $[0.7, 0.99]$, $[0.99, 0.996]$, and $[0.996, 1.0]$.

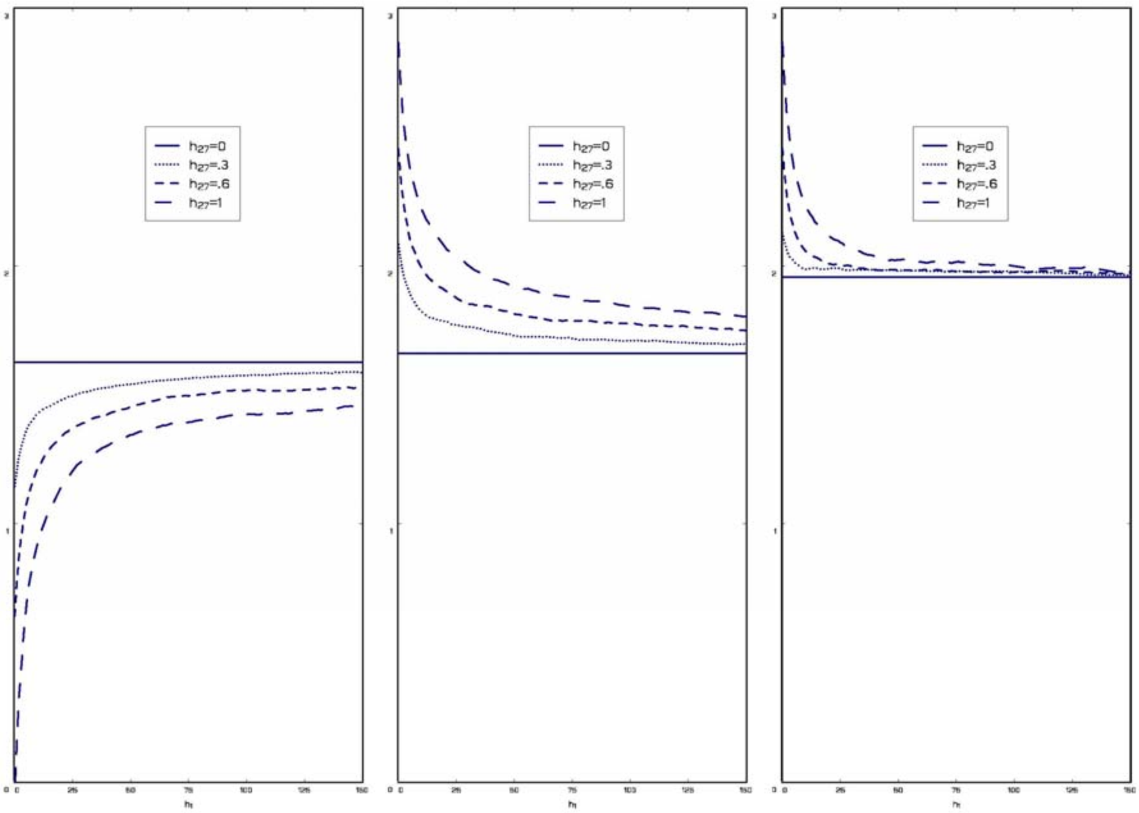


FIGURE 1.— Autoregression Example, Model 1: .95 Quantile Graphs, $c_h(.95)$, for J_h^*, $-J_h^*$, and $|J_h^*|$ (left, center, and right panel) as Functions of h_1 for Several Values of h_{27}

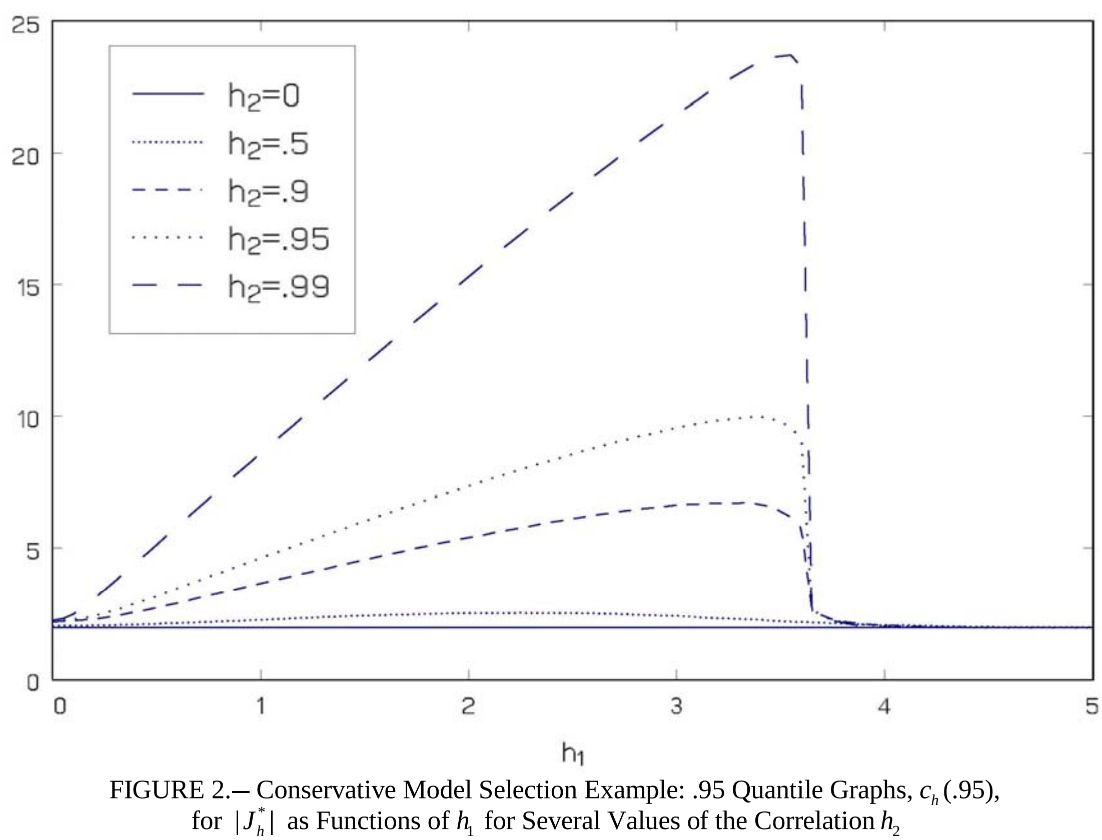


FIGURE 2.— Conservative Model Selection Example: .95 Quantile Graphs, $c_h(.95)$, for $|J_h^*|$ as Functions of h_1 for Several Values of the Correlation h_2

Supplement to
Hybrid and Size-Corrected Subsampling Methods

Donald W. K. Andrews*
Cowles Foundation for Research in Economics
Yale University

Patrik Guggenberger
Department of Economics
UCLA

June 2005
Revised: October 2008

* Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-0417911 and SES-0751517. Guggenberger gratefully acknowledges research support from a faculty research grant from UCLA in 2005 and from the National Science Foundation via grant number SES-0748922.

Abstract

This Supplement to "Hybrid and Size-Corrected Subsampling Methods" contains 11 sections of results. Section 1 provides details concerning Tables II and III. Section 2 gives the proofs of the results in the paper. Section 3 introduces size-correction methods based on quantile adjustment. Section 4 provides results concerning power comparisons of size-corrected (SC) tests. Section 5 provides graphical illustrations of the critical value functions of fixed critical value (FCV), subsampling, and hybrid tests. Section 6 gives graphical illustrations of power comparisons of SC-FCV, SC-Sub, and SC-Hyb tests. Section 7 introduces and give results for equal-tailed size-corrected tests. Section 8 defines a size-corrected combined (SC-Com) test that combines the SC-Sub and SC-Hyb tests. Section 9 gives asymptotic and finite-sample results for hybrid, SC, and PSC tests for the Nuisance Parameter Near a Boundary Example of Andrews and Guggenberger (2009a), hereafter AG1. Section 10 provides a table of asymptotic and finite-sample results for upper and lower one-sided confidence intervals in the Autoregressive Parameter Example considered in the paper. Section 10 also verifies the assumptions for that example. Section 11 verifies the assumptions for the Conservative Model Selection Example considered in the paper.

1 Details Concerning Tables II and III

To implement the Kristensen and Linton (2006) estimator used in the results of Table II, we use two Newton-Raphson iterations, see their equation (17), and to initialize the iteration we use their closed form estimator, see p.326, in particular their equation (10), implemented with $w_1 = w_2 = w_3 = 1/3$ and with their $\widehat{\phi}$ Winsorized to the interval [.001,.999]. In each iteration step, we initialize the $\widehat{\sigma}^2_{k,t}$ (p.329, line 5↑) by setting it equal to the squared first data observation. For simplicity, this estimator has not been discretized and the GARCH(1, 1) process has not been truncated to conform to the theoretical results given in the Section 3.4 of Andrews and Guggenberger (2008) for the asymptotic equivalence of feasible and infeasible QGLS statistics. The subsample statistics use the full-sample estimator of the conditional heteroskedasticity $\{\widehat{\phi}_{n,i} : i \leq n\}$, which is justified because feasible and infeasible QGLS test statistics are asymptotically equivalent in the full sample and in subsamples.

In Table II, the parameter space for ρ is taken to be $[-0.9, 1.0]$ to minimize the effect of the choice of the lower bound on the FS-Min values of the subsampling and hybrid CIs because in most practical applications in economics, the parameter interval $(-1.0, -0.9]$ is not of interest. The effects are small. For the parameter spaces $[-0.999, 1.0]$ and $[-.9, 1.0]$, the respective FS-Min values of the symmetric subsampling CIs are 94.6 and 95.0 for case (i), 95.1 and 95.4 for case (ii), 92.8 and 94.6 for case (iv), and 95.6 and 95.8 for case (v). For the symmetric hybrid CIs, they are 95.9 and 96.0 for case (iii), 93.7 and 94.6 for case (iv), and 96.0 and 96.1 for case (v). For the equal-tailed hybrid CI, they are 93.1 and 93.5 for case (iv). No other results are affected by the choice of the lower bound of the parameter space.

The 119 subsamples used in Table III include 10 "wrap-around" subsamples that contain observations at the end and beginning of the sample, for example, observations indexed by $(110, ..., 120, 1)$. The choice of $q_n = 119$ subsamples is made because this reduces rounding errors when q_n is small when computing the sample quantiles of the subsample statistics. The values ν_α that solve $\nu_\alpha/(q_n + 1) = \alpha$ for $\alpha = .025$, .95, and .975 are the integers 3, 114, and 117. In consequence, the .025, .95, and .975 sample quantiles are given by the 3rd, 114th, and 117th largest subsample statistics. See Hall (1992, p. 307) for a discussion of this choice in the context of the bootstrap.

2 Proofs

For notational simplicity, throughout this section, we let c_g, c_h, c_∞, $c_{n,b}$, and cv abbreviate $c_g(1-\alpha)$, $c_h(1-\alpha)$, $c_\infty(1-\alpha)$, $c_{n,b}(1-\alpha)$, and $cv(1-\alpha)$, respectively.

2.1 Proof of Lemma 1

Lemma 1 (of the Paper). *Suppose Assumptions* A-G, K, *and* T *hold. Then, either* (i) *the addition of* $c_\infty(1-\alpha)$ *to the subsampling critical value is irrelevant asymptotically* (*i.e.,* $c_h(1-\alpha) \geq c_\infty(1-\alpha)$ *for all* $h \in H$ and $Max_{Hyb}(\alpha) = Max_{Sub}(\alpha)$), *or* (ii) *the nominal level* α *subsampling test over-rejects asymptotically* (*i.e.,* $AsySz(\theta_0) > \alpha$) *and the hybrid test reduces the asymptotic over-rejection for at least one parameter value* $(g, h) \in GH$.

Proof of Lemma 1 (of the Paper). If $c_h \geq c_\infty$ for all $h \in H$, then $Max_{Hyb}(\alpha) = Max_{Sub}(\alpha)$ and $Max^-_{Hyb}(\alpha) = Max^-_{Sub}(\alpha)$ follow immediately (where the latter three quantities are defined in Assumptions P and T). In addition, Assumption T implies that all of these quantities are equal. The latter, Theorem 1 of the paper, and Theorem 1(ii) of AG1 imply that the quantities equal $AsySz(\theta_0)$ for the hybrid and subsampling tests.

On the other hand, suppose "$c_h \geq c_\infty$ for all $h \in H$" does not hold. Then, for some $g \in H$, $c_g < c_\infty$. Given g, define $h_1 = (h_{1,1}, ..., h_{1,p})' \in H_1$ by $h_{1,m} = +\infty$ if $g_{1,m} > 0$, $h_{1,m} = -\infty$ if $g_{1,m} < 0$, $h_{1,m} = +\infty$ or $-\infty$ (chosen so that $(g, h) \in GH$) if $g_{1,m} = 0$ for $m = 1, ..., p$, and define $h_2 = g_2$. Let $h = (h_1, h_2)$. By construction, $(g, h) \in GH$. By Assumption K, $c_h = c_\infty$. Hence, we have

$$Max_{Sub}(\alpha) \geq 1 - J_h(c_g) > \alpha, \tag{2.1}$$

where the second inequality holds because $c_g < c_\infty = c_h$ and c_h is the infimum of values x such that $J_h(x) \geq 1 - \alpha$ or, equivalently, $1 - J_h(x) \leq \alpha$. Equation (2.1) and Theorem 1(ii) of AG1 imply that $AsySz(\theta_0) > \alpha$ for the subsampling test. The hybrid test reduces the asymptotic over-rejection of the subsampling test at (g, h) from being at least $1 - J_h(c_g) > \alpha$ to being at most $1 - J_h(c_\infty) = 1 - J_h(c_h) \leq \alpha$ (with equality if $J_h(\cdot)$ is continuous at c_h). $\square$

2.2 Proof of Lemma 2

Lemma 2 (of the Paper). *Suppose Assumptions* A-G, K, T, *and* Quant *hold. Then, the hybrid test based on $T_n(\theta_0)$ has $AsySz(\theta_0) = \alpha$.*

Proof of Lemma 2 (of the Paper). Suppose Assumption Quant (i) holds. Then,

$$\begin{aligned} Max_{Hyb}(\alpha) &= \sup_{(g,h) \in GH} [1 - J_h(\max\{c_g, c_\infty\})] = \sup_{h \in H} [1 - J_h(c_\infty)] \\ &\leq \sup_{h \in H} [1 - J_h(c_h)] = \alpha, \end{aligned} \tag{2.2}$$

where the second equality and the inequality hold by Assumption Quant (i)(a) and the last equality holds because $1 - J_h(c_h) \leq \alpha$ by definition of c_h for all $h \in H$ and $1 - J_\infty(c_\infty) = \alpha$ by Assumption Quant (i)(b). By (2.2) and Assumption Quant (i)(b), $Max_{Hyb}(\alpha) = \sup_{h \in H}[1 - J_h(c_\infty)] \geq 1 - J_\infty(c_\infty) = \alpha$.

Next, suppose Assumption Quant (ii) holds. By Assumption Quant (ii)(a), $p = 1$. Hence, given $(g, h) \in GH$ either (I) $|h_{1,1}| = \infty$ or (II) $|h_{1,1}| < \infty$. When (I) holds, $J_h = J_\infty$ by Assumption K and

$$1 - J_h(\max\{c_g, c_\infty\}) \leq 1 - J_\infty(c_\infty) = \alpha. \tag{2.3}$$

When (II) holds, g must equal h^0 by the definition of GH. Hence,

$$1 - J_h(\max\{c_g, c_\infty\}) \leq 1 - J_h(c_{h^0}) \leq \sup_{h \in H} [1 - J_h(c_h)] = \alpha, \tag{2.4}$$

where the second inequality holds because $c_{h^0} \geq c_h$ by Assumption Quant (ii)(b) and the equality holds by Assumption Quant (ii)(c). Hence, $Max_{Hyb}(\alpha) \leq \alpha$. In addition, $Max_{Hyb}(\alpha) \geq 1 - J_\infty(c_\infty) = \alpha$ by Assumption Quant (ii)(c). $\square$

2.3 Proof of Theorem 2

In this section, we prove Theorem 2 of the paper. For the reader's convenience, we repeat the definition of the size-corrected (SC) tests here. The size-corrected fixed critical value (SC-FCV), subsampling (SC-Sub), and hybrid (SC-Hyb) tests with nominal level α are defined to reject the null hypothesis $H_0 : \theta = \theta_0$ when

$$\begin{aligned} &T_n(\theta_0) > cv(1 - \alpha), \\ &T_n(\theta_0) > c_{n,b}(1 - \alpha) + \kappa(\alpha) \text{ and} \\ &T_n(\theta_0) > \max\{c_{n,b}(1 - \alpha), c_\infty(1 - \alpha) + \kappa^*(\alpha)\}, \end{aligned} \tag{2.5}$$

respectively, where

$$
\begin{aligned}
cv(1-\alpha) &= \sup_{h\in H} c_h(1-\alpha), \\
\kappa(\alpha) &= \sup_{(g,h)\in GH} [c_h(1-\alpha) - c_g(1-\alpha)], \\
\kappa^*(\alpha) &= \sup_{h\in H^*} c_h(1-\alpha) - c_\infty(1-\alpha), \text{ and} \\
H^* &= \{h \in H : \text{for some } (g,h) \in GH, c_g(1-\alpha) < c_h(1-\alpha)\}.
\end{aligned}
\tag{2.6}
$$

If H^* is empty, then $\kappa^*(\alpha) = -\infty$ by definition.

Theorem 2 (of the Paper). *Suppose Assumptions* A-G *and* K-M *hold. Then, the* SC-FCV, SC-Sub, *and* SC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.

Proof of Theorem 2 (of the Paper). First we note that Assumption L implies that cv, $\kappa(\alpha)$, and $\kappa^*(\alpha)$ are finite. Below we show that cv, $\kappa(\alpha)$, and $\kappa^*(\alpha)$ satisfy

$$
\begin{aligned}
\sup_{h\in H}[1 - J_h(cv-)] &\leq \alpha, \\
\sup_{(g,h)\in GH} (1 - J_h((c_g + \kappa(\alpha))-)) &\leq \alpha, \text{ and} \\
\sup_{(g,h)\in GH} (1 - J_h(\max\{c_g, c_\infty + \kappa^*(\alpha)\}-)) &\leq \alpha,
\end{aligned}
\tag{2.7}
$$

respectively. Given (2.7), Theorem 1(i) of AG1 applied with $c_{Fix} = cv$ implies that the SC-FCV test satisfies $AsySz(\theta_0) \leq \sup_{h\in H}[1 - J_h(cv-)] \leq \alpha$, where the second inequality holds by (2.7). Theorem 1(ii) of AG1 with $c_{n,b}+\kappa(\alpha)$ in place of $c_{n,b}$ implies that the SC-Sub test satisfies $AsySz(\theta_0) \leq \sup_{(g,h)\in H}[1-J_h((c_g+\kappa(\alpha))-)] \leq \alpha$, where the second inequality holds by (2.7). Theorem 1(ii) of AG1 with $\max\{c_{n,b}, c_\infty+\kappa^*(\alpha)\}$ in place of $c_{n,b}$ implies that the SC-Hyb test satisfies $AsySz(\theta_0) \leq \sup_{(g,h)\in H}[1 - J_h(\max\{c_g, c_\infty + \kappa^*(\alpha)\}-)] \leq \alpha$, where the second inequality holds by (2.7). Hence, $AsySz(\theta_0) \leq \alpha$ for SC-FCV, SC-Sub, and SC-Hyb tests. Below we show that the reverse inequality also holds.

We now show that the first inequality in (2.7) holds. For $h \in H$, if $c_h < \sup_{h^\dagger\in H} c_{h^\dagger}$, then

$$
J_h\left(\sup_{h^\dagger\in H} c_{h^\dagger}-\right) \geq J_h(c_h) \geq 1-\alpha, \tag{2.8}
$$

where the first inequality holds because J_h is nondecreasing and the second inequality holds by the definition of c_h. For $h \in H$, if $c_h = \sup_{h^\dagger\in H} c_{h^\dagger}$, then

$$
J_h\left(\sup_{h^\dagger\in H} c_{h^\dagger}-\right) = J_h(c_h-) = 1-\alpha, \tag{2.9}
$$

where the last equality holds by Assumption M(a)(ii). For cv defined in (2.6), (2.8) and (2.9) combine to give

$$
\sup_{h\in H}[1 - J_h(cv-)] = \sup_{h\in H}[1 - J_h(\sup_{h^\dagger\in H} c_{h^\dagger}-)] \leq \alpha. \tag{2.10}
$$

Hence, cv satisfies (2.7).

Next, we prove that the second inequality in (2.7) holds. For $(g,h) \in GH$, if $c_h < c_g + \sup_{(g^\dagger,h^\dagger)\in GH}[c_{h^\dagger} - c_{g^\dagger}]$, then we have

$$J_h\left((c_g + \kappa(\alpha))-\right) = J_h\left((c_g + \sup_{(g^\dagger,h^\dagger)\in GH}[c_{h^\dagger} - c_{g^\dagger}])-\right) \geq J_h\left(c_h\right) \geq 1-\alpha, \quad (2.11)$$

where the first inequality holds by the condition on (g,h) and the fact that J_h is nondecreasing.

For $(g,h) \in GH$, if $c_h = c_g + \sup_{(g^\dagger,h^\dagger)\in GH}[c_{h^\dagger} - c_{g^\dagger}]$, then we have

$$J_h\left((c_g + \kappa(\alpha))-\right) = J_h\left((c_g + \sup_{(g^\dagger,h^\dagger)\in GH}[c_{h^\dagger} - c_{g^\dagger}])-\right) = J_h\left(c_h-\right) = 1-\alpha, \quad (2.12)$$

where the second equality holds by the condition on (g,h) and the last equality holds by Assumption M(b)(ii). Combining (2.11) and (2.12) gives $\sup_{(g,h)\in GH}[1 - J_h((c_g + \kappa(\alpha))-)] \leq \alpha$, as desired.

The third inequality in (2.7) holds by the following argument. Because $c_\infty + \kappa^*(\alpha) = \sup_{h^*\in H^*} c_{h^*}$, we need to show that $\sup_{(g,h)\in GH}[1-J_h(\max\{c_g, \sup_{h^*\in H^*} c_{h^*}\}-)] \leq \alpha$. For all $(g,h) \in GH$, we have $\max\{c_g, \sup_{h^*\in H^*} c_{h^*}\} \geq c_h$ because $\max\{c_g, \sup_{h^*\in H^*} c_{h^*}\} < c_h$ implies that $c_g < c_h$, which implies that $h \in H^*$, which implies that $\sup_{h^*\in H^*} c_{h^*} \geq c_h$, which is a contradiction. Now, for any $(g,h) \in GH$ with $\max\{c_g, \sup_{h^*\in H^*} c_{h^*}\} > c_h$, we have $1-J_h(\max\{c_g, \sup_{h^*\in H^*} c_{h^*}\}-) \leq 1-J_h(c_h) \leq \alpha$, as desired. For any $(g,h) \in GH$ with $\max\{c_g, \sup_{h^*\in H^*} c_{h^*}\} = c_h$, Assumption M(c)(ii) implies that $J_h(x)$ is continuous at $x = c_h$. Hence, $1 - J_h(\max\{c_g, c_\infty + \kappa^*(\alpha)\}-) = 1 - J_h(c_h-) = 1 - J_h(c_h) = \alpha$, which completes the proof of the third inequality of (2.7). This concludes the proof that $AsySz(\theta_0) \leq \alpha$ for the SC-FCV, SC-Sub, and SC-Hyb tests.

We now prove that these tests satisfy $AsySz(\theta_0) \geq \alpha$. By Theorem 1(i) of AG1 applied with $c_{Fix} = cv$, the SC-FCV test satisfies $AsySz(\theta_0) \geq \sup_{h\in H}[1 - J_h(cv)]$. Using (2.6) and Assumption M(a)(i), $cv = \sup_{h\in H} c_h = c_{h^*}$ for some $h^* \in H$. Hence,

$$\sup_{h\in H}[1 - J_h(cv)] = \sup_{h\in H}[1 - J_h(c_{h^*})] \geq 1 - J_{h^*}(c_{h^*}) = \alpha, \quad (2.13)$$

where the last equality holds by Assumption M(a)(ii). In consequence, for the SC-FCV test, $AsySz(\theta_0) \geq \alpha$.

Next, by Theorem 1(ii) of AG1 with $c_{n,b} + \kappa(\alpha)$ in place of $c_{n,b}$, the SC-Sub test satisfies $AsySz(\theta_0) \geq \sup_{(g,h)\in GH}[1 - J_h(c_g + \kappa(\alpha))]$. Using (2.6) and Assumption M(b)(i), $\kappa(\alpha) = c_{h^*} - c_{g^*}$ for some $(g^*,h^*) \in GH$ as in Assumption M(b)(i). Hence,

$$\sup_{(g,h)\in GH}[1-J_h(c_g+\kappa(\alpha))] = \sup_{(g,h)\in GH}[1-J_h(c_g+c_{h^*}-c_{g^*})] \geq 1-J_{h^*}(c_{h^*}) = \alpha, \quad (2.14)$$

where the last equality holds by Assumption M(b)(ii). In consequence, for the SC-Sub test, $AsySz(\theta_0) \geq \alpha$.

Lastly, Theorem 1(ii) of AG1 with $\max\{c_{n,b}, c_\infty + \kappa^*(\alpha)\}$ in place of $c_{n,b}$ implies that the SC-Hyb test satisfies

$$AsySz(\theta_0) \geq \sup_{(g,h)\in GH} [1 - J_h(\max\{c_g, c_\infty + \kappa^*(\alpha)\})]. \tag{2.15}$$

If H^* is not empty, then using (2.6) and Assumption M(c)(i), $\kappa^*(\alpha) = c_{h^*} - c_\infty$ for some $h^* \in H^*$ as in Assumption M(c)(i). By the definition of H^*, there exists g^* such that $(g^*, h^*) \in GH$ and $c_{g^*} < c_{h^*}$. In consequence, the right-hand side of (2.15) equals

$$\sup_{(g,h)\in GH} [1 - J_h(\max\{c_g, c_{h^*}\})] \geq 1 - J_{h^*}(\max\{c_{g^*}, c_{h^*}\}) = 1 - J_{h^*}(c_{h^*}) = \alpha, \tag{2.16}$$

where the first equality uses $c_{g^*} < c_{h^*}$ and the last equality holds by Assumption M(c)(ii) because $(g^*, h^*) \in GH$ satisfies $c_{h^*} = \sup_{h\in H^*} c_h = \max\{c_{g^*}, \sup_{h\in H^*} c_h\}$. Combining (2.15) and (2.16) gives $AsySz(\theta_0) \geq \alpha$.

If H^* is empty, then $\kappa^*(\alpha) = -\infty$, $(h^0, h^0) \in GH$, where $h^0 = (0, h_2)$ for arbitrary $h_2 \in H_2$, and we have

$$\begin{aligned} &\sup_{(g,h)\in GH} [1 - J_h(\max\{c_g, c_\infty + \kappa^*(\alpha)\})] \\ = &\sup_{(g,h)\in GH} [1 - J_h(c_g)] \geq 1 - J_{h^0}(c_{h^0}) = \alpha, \end{aligned} \tag{2.17}$$

where the last equality holds by Assumption M(c)(ii) because $c_{h^0} = \max\{c_{h^0}, c_\infty + \kappa^*(\alpha)\}$. Combining (2.15)–(2.17) gives $AsySz(\theta_0) \geq \alpha$ for the SC-Hyb test. □

2.4 Proof of Theorem 3

Theorem 3 (of the Paper). *Suppose Assumptions* A-G, K, L, N, *and* O *hold. Then,* (a) $cv_{\widehat{\gamma}_{n,2}}(1-\alpha) - cv_{\gamma_{n,2}}(1-\alpha) \to_p 0$, $\kappa_{\widehat{\gamma}_{n,2}}(\alpha) - \kappa_{\gamma_{n,2}}(\alpha) \to_p 0$, *and* $\kappa^*_{\widehat{\gamma}_{n,2}}(\alpha) - \kappa^*_{\gamma_{n,2}}(\alpha) \to_p 0$ *under all sequences* $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\}$, *and* (b) *the* PSC-FCV, PSC-Sub, *and* PSC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.

Proof of Theorem 3 (of the Paper). The results of part (a) hold by an extension of Slutsky's Theorem (to allow $\gamma_{n,2}$ to depend on n) using Assumption N and the uniform continuity of the functions in Assumptions O(a)(i), O(b)(i), and O(c)(i). The proof of part (b) is split into two steps. In the first step, we consider the PSC tests with $\widehat{\gamma}_{n,2}$ replaced by the true value $\gamma_{n,2}$. In this case, using parts (ii) and (iii) of Assumptions O(a), O(b), and O(c), the results of part (b) hold by a very similar argument to that given in the proof of Theorem 2 of the paper. In the second step, the results of parts (a) are combined with the results of the first step to obtain the desired results. This step holds because the results of parts (a) lead to the same limit distributions for the statistics in question whether they are based on $\widehat{\gamma}_{n,2}$ or the true value $\gamma_{n,2}$ by the argument used in the proof of Theorem 1(ii) of AG1. □

2.5 Proof of Theorem 4

Theorem 4(a) (of the Paper). *Suppose Assumptions* A-G *and* P *hold. Then, a subsampling test satisfies*

$$\lim_{n\to\infty} AsySz_n(\theta_0) = AsySz(\theta_0).$$

Proof of Theorem 4(a) (of the Paper). Under Assumptions A-G, Theorem 1 of AG1 combined with Assumption P(ii) shows that $AsySz(\theta_0) = \sup_{(g,h)\in GH}(1 - J_h(c_g))$. First, we show that $\liminf_{n\to\infty} AsySz_n(\theta_0) \geq AsySz(\theta_0)$. Given $(g,h) = ((g_1,h_2),(h_1,h_2)) \in GH$, we construct a sequence $\{h_n = (h_{n,1},h_{n,2}) \in H : n \geq 1\}$ such that $(g_n,h_n) \to (g,h)$ as $n\to\infty$, where $g_n = (g_{n,1},g_{n,2}) = (\delta_n^r h_{n,1}, h_{n,2})$. Define $h_{n,2} = h_2$ for all $n \geq 1$. We write $h_1 = (h_{1,1},...,h_{1,p})'$ and $h_{n,1} = (h_{n,1,1},...,h_{n,1,p})'$. For $m = 1,...,p$, define

$$\begin{array}{ll} h_{n,1,m} = h_{1,m} & \text{if } g_{1,m} = 0 \ \&\ |h_{1,m}| < \infty \\ h_{n,1,m} = (n/b_n)^{r/2} & \text{if } g_{1,m} = 0 \ \&\ h_{1,m} = \infty \\ h_{n,1,m} = -(n/b_n)^{r/2} & \text{if } g_{1,m} = 0 \ \&\ h_{1,m} = -\infty \\ h_{n,1,m} = (n/b_n)^r g_{1,m} & \text{if } g_{1,m} \in (0,\infty) \ \&\ h_{1,m} = \infty \\ h_{n,1,m} = (n/b_n)^r g_{1,m} & \text{if } g_{1,m} \in (-\infty,0) \ \&\ h_{1,m} = -\infty \\ h_{n,1,m} = (n/b_n)^{2r} & \text{if } g_{1,m} = \infty \ \&\ h_{1,m} = \infty \\ h_{n,1,m} = -(n/b_n)^{2r} & \text{if } g_{1,m} = -\infty \ \&\ h_{1,m} = -\infty. \end{array} \quad (2.18)$$

As defined, $(g_{n,1},h_{n,1}) = (\delta_n^r h_{n,1}, h_{n,1}) \to (g_1,h_1)$ and $(g_n,h_n) \to (g,h)$.

We now have

$$\begin{aligned} \liminf_{n\to\infty} AsySz_n(\theta_0) &= \liminf_{n\to\infty} \sup_{h=(h_1,h_2)\in H} (1 - J_h(c_{(\delta_n^r h_1,h_2)})) \\ &\geq \liminf_{n\to\infty} (1 - J_{h_n}(c_{(\delta_n^r h_{n,1},h_{n,2})})) \\ &= \liminf_{n\to\infty} (1 - J_{h_n}(c_{g_n})) \\ &= 1 - J_h(c_g), \end{aligned} \quad (2.19)$$

where the second equality holds by definition of g_n and the last equality holds by Assumption P because $(g_n,h_n) \to (g,h)$. Using the expression for $AsySz(\theta_0)$ given above, this establishes the desired result because (2.19) holds for all $(g,h) \in GH$.

Next, we show that $\limsup_{n\to\infty} AsySz_n(\theta_0) \leq AsySz(\theta_0)$. For $h = (h_1,h_2) \in H$, let $\tau_n(h) = 1 - J_h(c_{(\delta_n^r h_1,h_2)})$. By definition, $AsySz_n(\theta_0) = \sup_{h\in H}\tau_n(h)$. There exists a sequence $\{h_n \in H : n \geq 1\}$ such that $\limsup_{n\to\infty} \sup_{h\in H}\tau_n(h) = \limsup_{n\to\infty} \tau_n(h_n)$. There exists a subsequence $\{u_n\}$ of $\{n\}$ such that $\limsup_{n\to\infty}\tau_n(h_n) = \lim_{n\to\infty}\tau_{u_n}(h_{u_n})$. There exists a subsequence $\{v_n\}$ of $\{u_n\}$ such that $(h_{v_n,1}, h_{v_n,2}, \delta_{v_n}^r h_{v_n,1}) \to (h_1^*, h_2^*, g_1^*)$ for some $h_1^* \in H_1$, $h_2^* \in H_2$, $g_1^* \in H_1$, where $(g^*,h^*) = ((g_1^*,h_2^*),(h_1^*,h_2^*)) \in GH$. Hence,

$$\limsup_{n\to\infty} AsySz_n(\theta_0) = \lim_{n\to\infty}\tau_{u_n}(h_{u_n}) = \lim_{n\to\infty}\tau_{v_n}(h_{v_n})$$

$$= \lim_{n\to\infty} \left(1 - J_{h_{v_n}}(c_{(\delta_{v_n}^r h_{v_n,1}, h_{v_n,2})})\right) = 1 - J_{h^*}(c_{g^*})$$
$$\leq \sup_{(g,h)\in GH} (1 - J_h(c_g)) = AsySz(\theta_0), \tag{2.20}$$

where the fourth equality holds by Assumption P and the results above. □

Theorem 4(b) (of the Paper). *Suppose Assumptions* A-G, K-M, Q, *and* R *hold. Then,*

(i) $\lim_{n\to\infty} \kappa(\delta_n, \alpha) = \kappa(\alpha)$ *and* $\lim_{n\to\infty} \kappa^*(\delta_n, \alpha) = \kappa^*(\alpha)$ *and*
(ii) *the* ASC-Sub *and* ASC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.

Proof of Theorem 4(b) (of the Paper). The first result of part (i) holds by the proof of Theorem 4(a) of the paper with $1 - J_h(c_{(\delta_n^r h_1, h_2)})$ and $1 - J_h(c_g)$ replaced by $c_{(h_1,h_2)} - c_{(\delta_n^r h_1, h_2)}$ and $c_h - c_g$, respectively, using Assumption Q in place of Assumption P. Next, we show the first result of part (ii). Using the first result of part (i), by the same argument as used to prove Theorem 1(ii) of AG1, $AsySz(\theta_0)$ for the ASC-Sub test equals $AsySz(\theta_0)$ for the SC-Sub test. By Theorem 2 of the paper, the latter equals α.

Now, we prove that the second result of part (i) holds with $\lim_{n\to\infty}$ and $=$ replaced by $\liminf_{n\to\infty}$ and $\geq$, respectively, even without imposing Assumption R. If H^* is empty, then $\liminf_{n\to\infty} \sup_{h\in H^*(\delta_n)} c_h \geq -\infty = \sup_{h\in H^*} c_h$. If H^* is non-empty, for any $(g, h) \in GH$ such that $h \in H^*$, define $(g_n, h_n) \in GH$ as in (2.18). By $(g_n, h_n) \to (g, h)$, Assumption Q, and $c_g - c_h < 0$, we obtain $c_{g_n} - c_{h_n} < 0$ and $h_n \in H^*(\delta_n)$ for all n sufficiently large. Hence,

$$\liminf_{n\to\infty} \sup_{h\in H^*(\delta_n)} c_h \geq \liminf_{n\to\infty} c_{h_n} = c_h, \tag{2.21}$$

where the equality uses $h_n \to h$ and Assumption Q. This inequality holds for all $h \in H^*$. Hence, $\liminf_{n\to\infty} \sup_{h\in H^*(\delta_n)} c_h \geq \sup_{h\in H^*} c_h$ and the proof is complete.

Next, we show the second result of part (ii) holds with $=$ replaced by $\leq$ even without imposing Assumption R. Using the second result of part (i) with $\lim_{n\to\infty}$ and $=$ replaced by $\liminf_{n\to\infty}$ and $\geq$, respectively, the $\limsup_{n\to\infty}$ of the rejection probability of the ASC-Hyb test is less than or equal to that of the SC-Hyb test and the latter equals α by Theorem 1 of the paper.

To show that the second result of part (i) holds, it remains to show that it holds with $=$ replaced by $\leq$. First suppose that H^* is empty. Then, $\kappa^*(\alpha) = -\infty$, $H^*(\delta)$ is empty for $\delta > 0$ close to zero by Assumption R, and $\kappa^*(\delta_n, \alpha) = -\infty$ for n sufficiently large. Next, suppose that H^* is non-empty. Then, using Assumption R, it suffices to show that $\limsup_{n\to\infty} \sup_{h\in H^*(\delta_n)} c_h \leq \sup_{h\in H^\dagger} c_h$. As in the last paragraph of the proof of Theorem 4(a) of the paper (given above), there exists a sequence $\{h_n \in H^*(\delta_n) : n \geq 1\}$, a subsequence $\{u_n\}$ of $\{n\}$, and a subsequence $\{v_n\}$ of $\{u_n\}$ such that

$$\limsup_{n\to\infty} \sup_{h\in H^*(\delta_n)} c_h = \lim_{n\to\infty} c_{h_{v_n}}, \tag{2.22}$$

$$(h_{v_n,1}, h_{v_n,2}) \to (h_1^*, h_2^*) = h^*, \text{ and } (\delta_{v_n}^r h_{v_n,1}, h_{v_n,2}) \to (g_1^*, h_2^*) = g^*$$

for some $(g^*, h^*) \in GH$. Since $h_{v_n} = (h_{v_n,1}, h_{v_n,2}) \in H^*(\delta_{v_n})$ for all n, we have $h^* \in H^\dagger$ by definition of $H^\dagger$. This, (2.22), and Assumption Q yield

$$\limsup_{n\to\infty} \sup_{h\in H^*(\delta_n)} c_h = \lim_{n\to\infty} c_{h_{v_n}} = c_{h^*} \leq \sup_{h\in H^\dagger} c_h, \tag{2.23}$$

which completes the proof of the second result of part (i). Given this, by the same argument as used to prove Theorem 1(ii) of AG1 with $c_{n,b}$ replaced by $\max\{c_{n,b}, c_\infty + \kappa^*(\delta_n, \alpha)\}$, $AsySz(\theta_0)$ for the ASC-Hyb test is equal to $AsySz(\theta_0)$ for the SC-Hyb test. By Theorem 2, the latter equals α. Hence, the second result of part (ii) holds. $\square$

Theorem 4(c) (of the Paper). *Suppose Assumptions* A-G, K, L, N, O, Q, *and* S *hold. Then,*

(i) $\kappa_{\widehat{\gamma}_{n,2}}(\delta_n, \alpha) - \kappa_{\gamma_{n,2}}(\alpha) \to_p 0$ *and* $\kappa^*_{\widehat{\gamma}_{n,2}}(\delta_n, \alpha) - \kappa^*_{\gamma_{n,2}}(\alpha) \to_p 0$ *under all sequences* $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\}$ *and*

(ii) *the* APSC-Sub *and* APSC-Hyb *tests satisfy* $AsySz(\theta_0) = \alpha$.

Proof of Theorem 4(c) (of the Paper). By Theorem 3 of the paper, in part (a) it suffices to show that $\kappa_{\widehat{\gamma}_{n,2}}(\delta_n, \alpha) - \kappa_{\widehat{\gamma}_{n,2}}(\alpha) \to_p 0$ and $\kappa^*_{\widehat{\gamma}_{n,2}}(\delta_n, \alpha) - \kappa^*_{\widehat{\gamma}_{n,2}}(\alpha) \to_p 0$. To do so, we use the result that a sequence of random variables $\{X_n : n \geq 1\}$ satisfies $X_n \to_p 0$ if and only if for every subsequence $\{u_n\}$ of $\{n\}$ there is a subsequence $\{v_n\}$ of $\{u_n\}$ such that $X_{v_n} \to 0$ a.s. We apply this result with $X_n = \kappa_{\widehat{\gamma}_{n,2}}(\delta_n, \alpha) - \kappa_{\widehat{\gamma}_{n,2}}(\alpha)$. Hence, it suffices to show that given any $\{u_n\}$ there exists a subsequence $\{v_n\}$ of $\{u_n\}$ such that $X_{v_n} \to 0$ a.s. Given $\{u_n\}$, we apply the above subsequence result a second time with $X_n = \widehat{\gamma}_{n,2} - \gamma_{n,2}$ to guarantee that there is a subsequence $\{v_n\}$ of $\{u_n\}$ for which $\widehat{\gamma}_{v_n,2} - \gamma_{v_n,2} \to 0$ a.s. using Assumption N. The subsequence $\{v_n\}$ can be chosen such that $\gamma_{v_n,2} \to h_2$ for some $h_2 \in H_2$ because every sequence in H_2 has a convergent subsequence given that H_2 is closed with respect to R^q_∞. Now, the argument in the proof of Theorem 4(a) of the paper applied to the subsequence $\{v_n\}$ with $1 - J_h(c_{(\delta_{v_n}^r h_1, h_2)})$, $1 - J_h(c_g)$, and $h_{v_n,2} = h_2$ replaced by $c_{(h_1, \widehat{\gamma}_{v_n,2})} - c_{(\delta_{v_n}^r h_1, \widehat{\gamma}_{v_n,2})}$, $c_h - c_g$, and $h_{v_n,2} = \widehat{\gamma}_{v_n,2}$, respectively, and using Assumption Q in place of Assumption P, gives the desired result.

The second result of part (i) holds using similar subsequence arguments to those above combined with variations of the proof of the second result of part (i) of Theorem 4(b) of the paper with H^*, $H^*(\delta_n)$, $H^\dagger$, and Assumption R replaced by $H^*_{h_2}$, $H^*_{\gamma_{n,2}}(\delta_n)$, $H^\dagger_{h_2}$, and Assumption S, respectively.

Given the results of part (i), part (ii) is proved using the same argument as used to prove part (ii) of Theorem 4(b) of the paper. $\square$

3 Size Correction By Quantile Adjustment

We now briefly discuss SC methods based on quantile adjustment, as opposed to the method in Section 3 of the paper. Quantile-adjusted SC-Sub and SC-Hyb tests with nominal level α reject the null hypothesis $H_0 : \theta = \theta_0$ when

$$\begin{aligned} &T_n(\theta_0) > c_{n,b}(1 - \xi(\alpha)) \text{ and} \\ &T_n(\theta_0) > c^*_{n,b}(1 - \xi^*(\alpha)), \end{aligned} \tag{3.1}$$

respectively, where $\xi(\alpha)$ $(\in (0, \alpha])$, and $\xi^*(\alpha)$ $(\in (0, \alpha])$ are the largest constants[1] that satisfy

$$\begin{aligned} &\sup_{(g,h)\in GH} (1 - J_h(c_g(1 - \xi(\alpha))-)) \leq \alpha \text{ and} \\ &\sup_{(g,h)\in GH} (1 - J_h(\max\{c_g(1 - \xi^*(\alpha)), c_\infty(1 - \xi^*(\alpha))\}-)) \leq \alpha. \end{aligned} \tag{3.2}$$

In many cases, the quantile adjustment and the size-correction method of Section 3 of the paper give similar results. For many examples, we prefer the method based on (2.5)-(2.6) to that of (3.1)-(3.2) because the former are based on the explicit formulae for the adjustment factors $\kappa(\alpha)$ and $\kappa^*(\alpha)$ given in (2.6).

4 Power Comparisons of Size-Corrected Tests

We now provide some results concerning power comparisons of SC tests that are referred to in Section 3.2 of the paper. We consider three alternative assumptions concerning the shape of $c_h(1 - \alpha)$. ("Quant" refers to "quantile.")

Assumption Quant1. $c_g(1 - \alpha) \geq c_h(1 - \alpha)$ for all $(g, h) \in GH$.

Assumption Quant2. $c_g(1-\alpha) \leq c_h(1-\alpha)$ for all $(g, h) \in GH$ with strict inequality for some (g, h).

Assumption Quant3. (i) $H = H_1 = R_{+,\infty}$, (ii) $c_h(1 - \alpha)$ is uniquely maximized at $h^* \in (0, \infty)$, and (iii) $c_h(1 - \alpha)$ is minimized at $h = 0$ or $h = \infty$.

Theorem S1. *Suppose Assumptions* K *and* L *hold.*

(a) *Suppose Assumption* Quant1 *holds. Then,* (i) $cv(1 - \alpha) = \sup_{h_2\in H_2} c_{(0,h_2)}(1 - \alpha)$, (ii) $\kappa(\alpha) = 0$, (iii) $\kappa^*(\alpha) = -\infty$, (iv) $\max\{c_g(1 - \alpha), c_\infty(1 - \alpha) + \kappa^*(\alpha)\} = c_g(1 - \alpha) + \kappa(\alpha)$, and (v) $c_g(1 - \alpha) + \kappa(\alpha) \leq cv(1 - \alpha)$ for all $g \in H$.

(b) *Suppose Assumption* Quant2 *holds. Then,* (i) $cv(1 - \alpha) = c_\infty(1 - \alpha)$, (ii) $\kappa^*(\alpha) = 0$, (iii) $\max\{c_g(1 - \alpha), c_\infty(1 - \alpha) + \kappa^*(\alpha)\} = cv(1 - \alpha)$, *and* (iv) $cv(1 - \alpha) \leq c_g(1 - \alpha) + \kappa(\alpha)$ *for all* $g \in H$.

[1] If no such largest value exists, we take some value that is arbitrarily close to the supremum of the values that satisfy (3.2).

(c) *Suppose Assumption* Quant3 *holds. Then,* (i) $cv(1-\alpha) = c_{h^*}(1-\alpha)$, (ii) $\kappa(\alpha) = c_{h^*}(1-\alpha) - c_0(1-\alpha)$, (iii) $\kappa^*(\alpha) = c_{h^*}(1-\alpha) - c_\infty(1-\alpha)$, (iv) $\max\{c_g(1-\alpha), c_\infty(1-\alpha)+\kappa^*(\alpha)\} = cv(1-\alpha)$ for all $g \in H$, (v) $cv(1-\alpha) \leq c_g(1-\alpha)+\kappa(\alpha)$ for all $g \in H$ such that $c_g(1-\alpha) \geq c_0(1-\alpha)$ (*such as* $g = h^*$), *and likewise with strict inequalities, and* (vi) $cv(1-\alpha) > c_g(1-\alpha)+\kappa(\alpha)$ *for all* $g \in H$ *such that* $c_g(1-\alpha) < c_0(1-\alpha)$ (*there is no such* $g \in H$ *if* $c_h(1-\alpha)$ *is minimized at* $h = 0$).

Comments. 1. In this comment and the next, we assume Assumption M holds, so that Theorem 2 (of the paper) holds. Theorem S1(a) shows that the subsampling and hybrid tests have correct asymptotic size under Assumption Quant1 and they have critical values less than or equal to that of the SC-FCV test. Theorem S1(b) shows that the FCV and hybrid tests have correct asymptotic size under Assumption Quant2 and they have critical values less than or equal to that of the SC-Sub test. If Assumption Quant1 (Quant2) holds with a strict inequality for $(g,h) = (h^0, h)$ for some $h = (h_1, h_2) \in H$, where $h^0 = (0, h_2) \in H$, then Theorem S1(a)(v) (respectively, (b)(iv)) holds with a strict inequality with g equal to this value of h.

2. Theorem S1(c)(iv)-(v) shows that under Assumption Quant3 the SC-Hyb and SC-FCV tests are asymptotically equivalent and are always more powerful than the SC-Sub test at some $(g,h) \in GH$. On the other hand, Theorem S1(c)(vi) shows that under Assumption Quant3 the SC-Sub test can be more powerful than the SC-Hyb and SC-FCV tests at some $(g,h) \in GH$ though not if $c_h(1-\alpha)$ is minimized at $h = 0$.

The results above are relevant when the subsample statistics satisfy Assumption Sub1 (because then their asymptotic distribution typically is the same under the null and the alternative). On the other hand, if Assumption Sub2 holds, then the subsampling critical values typically diverge to infinity under fixed alternatives (at rate $b^{1/2} << n^{1/2}$ when $T_n(\theta_0)$ is a t statistic). For brevity, we do not investigate the relative magnitudes of the critical values of the SC-FCV, SC-Sub, and SC-Hyb tests for local alternatives when Assumption Sub2 holds.

In Section 8 below, we introduce a SC combined method that has power at least as good as that of the SC subsampling and hybrid tests. But, it reduces to the SC hybrid test in most examples and, hence, may be of more interest theoretically than practically.

Proof of Theorem S1. Assumption L guarantees that cv, $\kappa(\alpha)$, and $\kappa^*(\alpha)$ are well-defined. Part (a)(i) follows from the definition of cv in (2.6) and Assumption Quant1. Part (a)(ii) holds by definition of $\kappa(\alpha)$ in (2.6) and the fact that $c_h - c_g \leq 0$ for all $(g,h) \in GH$ by Assumption Quant1 (with equality for some $(g,h) \in GH$). Part (a)(iii) holds by the definition of $\kappa^*(\alpha)$ in (2.6) for the case where H^* is empty, because H^* is empty by Assumption Quant1. Part (a)(iv) follows from parts (a)(ii) and (a)(iii). Part (a)(v) follows from part (a)(ii) and the definition of cv in (2.6).

Next, we prove part (b)(i). Given any $g = (g_1, g_2) = (g_{1,1}, ..., g_{1,p}, g_2) \in H$, let $g^\infty = (g_1^\infty, g_2) = (g_{1,1}^\infty, ..., g_{1,p}^\infty, g_2) \in H$ be such that $g_{1,m}^\infty = +\infty$ if $g_{1,m} > 0$, $g_{1,m}^\infty = -\infty$

if $g_{1,m} < 0$, $g^{\infty}_{1,m} = +\infty$ or $-\infty$ (chosen so that $g^{\infty} \in H$) if $g_{1,m} = 0$ for $m = 1, ..., p$. By Assumption Quant2, $c_g \leq c_{g^{\infty}}$ because $(g, g^{\infty}) \in GH$. By Assumption K, $c_{g^{\infty}} = c_{\infty}$ for all $g \in H$. Hence, $cv = \sup_{h \in H} c_h = c_{\infty}$, which proves part (b)(i).

We now prove part (b)(ii). By Assumptions Quant2 and K, H^* is not empty and $\sup_{h \in H^*} c_h = c_{\infty}$. In consequence, $\kappa^*(\alpha) = 0$ by definition of $\kappa^*(\alpha)$ in (2.6). Part (b)(iii) follows from parts (b)(i) and (b)(ii) and $c_g \leq c_{\infty}$ by Assumptions Quant2 and K. We now prove part (b)(iv). By part (b)(i), it suffices to show that $c_g + \kappa(\alpha) \geq c_{\infty}$ for all $g \in H$. By the definition of $\kappa(\alpha)$ in (2.6) and Assumptions Quant2 and K, $\kappa(\alpha) = c_{\infty} - \inf_{h_2 \in H_2} c_{(0,h_2)}$. Hence, $c_g + \kappa(\alpha) = c_g + c_{\infty} - \inf_{h_2 \in H_2} c_{(0,h_2)} \geq c_{\infty}$, where the inequality uses Assumption Quant2. This establishes part (b)(iv).

Part (c)(i) holds by Assumption Quant3(ii). Part (c)(ii) holds by definition of $\kappa(\alpha)$ in (2.6) and Assumptions Quant3(ii) and Quant3(iii). Part (c)(iii) holds by definition of $\kappa^*(\alpha)$ in (2.6) and Assumption Quant3(ii). Part (c)(iv) holds because $\max\{c_g, c_{\infty} + \kappa^*(\alpha)\} = \max\{c_g, c_{h^*}\} = c_{h^*} = cv$ using parts (c)(i) and (c)(iii). Parts (c)(v) and (c)(vi) hold because $cv = c_{h^*}$ by part (c)(i) and $c_g + \kappa(\alpha) = c_{h^*} + c_g - c_0$ by part (c)(ii). $\square$

5 Critical Value Functions

In this section, we use graphs given in Figure B-1 to illustrate the asymptotic critical value (cv) functions of the hybrid, FCV, and subsampling tests for the case where $\gamma = \gamma_1 \in R_+$, (i.e., no subvectors γ_2 or γ_3 appear, $p = 1$, and $H = R_{+,\infty}$). The argument of the cv functions is $g \in H$. For example, the asymptotic subsampling cv function is $c_g(1-\alpha)$ for $g \in H$. In Figure B-1, the curved line is the subsampling cv function, the horizontal line is the FCV cv function, i.e., the constant $c_{\infty}(1-\alpha)$, and the hybrid cv function is the maximum of the two.

In Figure B-1(a), the subsampling and hybrid cv functions are the same and the corresponding tests have the desired asymptotic size α. (The latter holds because $c_{\infty}(1-\alpha)$ is $\leq$ the cv function at g for all $g \in R_+$, $c_0(1-\alpha)$ is $\geq$ the cv function at g for all $g \in R_+$, and these two conditions are necessary and sufficient for a test to have asymptotic size α assuming continuity of $J_h(\cdot)$ by Theorem 1 of AG1). On the other hand, in Figure B-1(a), the FCV test has asymptotic size $> \alpha$. In Figures B-1(b) and B-1(d), the hybrid cv function equals the FCV cv function, both of these tests have asymptotic size α, whereas the subsampling test has asymptotic size $> \alpha$. Figures B-1(a) and B-1(b) illustrate the results of Lemma 1(i) and 1(ii) of the paper, respectively.

Figure B-1(c) illustrates a case where the hybrid test has asymptotic size α, but both the FCV and subsampling tests have asymptotic size $> \alpha$. In Figures B-1(a)-(d), Assumption Quant holds, so the hybrid test has correct asymptotic size, as established in Lemma 1 of the paper.

Figures B-1(e) and B-1(f) illustrate cases in which the function $c_g(1-\alpha)$ is maximized at an interior point $g \in (0, \infty)$. In these cases, the hybrid, FCV, and subsampling tests all have asymptotic size $> \alpha$. Figures B-1(e) and B-1(f) illustrate the results of Lemma 1(ii) and 1(i) of the paper, respectively. In particular, in Figure B-1(e), the over-rejection of the subsampling test for g close to zero is reduced for the hybrid test because its cv function is larger.

6 Graphical Power Comparisons

Next, we use graphs given in Figure B-2 to illustrate the power comparison between SC-FCV, SC-Sub, and SC-Hyb tests. Theorem S1 shows that (a) if $c_g(1-\alpha) \geq c_h(1-\alpha)$ for all $(g, h) \in GH$, then the SC-Sub, SC-Hyb, Sub, and Hyb tests are equivalent asymptotically and are more powerful than the SC-FCV test, see Figure B-2(a); (b) if $c_g(1-\alpha) \leq c_h(1-\alpha)$ for all $(g, h) \in GH$, then the SC-FCV, SC-Hyb, FCV, and Hyb tests are equivalent asymptotically and are more powerful than the SC-Sub test, see Figure B-2(b); and (c) if $H = H_1 = R_{+,\infty}$ and $c_h(1-\alpha)$ is uniquely maximized at $h^* \in (0, \infty)$, then the SC-FCV and SC-Hyb tests are asymptotically equivalent and are either (i) more powerful than the SC-Sub test, see Figure B-2(e), or (ii) more powerful than the SC-Sub test for some values of $(g, h) \in GH$ but less powerful for other values of $(g, h) \in GH$, see Figure B-2(f).

Figure B-2(c) illustrates the case where $c_g(1-\alpha)$ is not monotone but is maximized at $g = 0$, the Hyb and SC-Hyb cv functions are the same, the Hyb cv function is lower than both the SC-Sub and SC-FCV cv functions, and so the Hyb test is more powerful than the SC-Sub and SC-FCV tests. Figure B-2(d) illustrates the case where $c_g(1-\alpha)$ is not monotone but is maximized at $g = \infty$, the Hyb, SC-Hyb, FCV, and SC-FCV cv functions are the same, the Hyb cv function is lower than the SC-Sub cv function, and so the Hyb test is more powerful than the SC-Sub test.

7 Equal-Tailed Size-Corrected Tests

This section introduces *equal-tailed* size-corrected FCV, subsampling, and hybrid t tests. It also introduces finite-sample-adjusted asymptotics for equal-tailed tests. We suppose that $T_n(\theta_0) = \tau_n(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n$.

Let $c_{Fix}(1-\alpha/2)$ and $c_{Fix}(\alpha/2)$ denote the critical values of the equal-tailed FCV test before size-correction. Equal-tailed (i) SC-FCV, (ii) SC-Sub, and (iii) SC-Hyb tests are defined by (5.1) of the paper with the critical values $c^*_{n,b}(1-\alpha/2)$ and $c^{**}_{n,b}(\alpha/2)$ replaced by (i) $c_{Fix}(1-\alpha/2) + \kappa_{ET,Fix}(\alpha)$ and $c_{Fix}(\alpha/2) - \kappa_{ET,Fix}(\alpha)$, (ii) $c_{n,b}(1-\alpha/2) + \kappa_{ET}(\alpha)$ and $c_{n,b}(\alpha/2) - \kappa_{ET}(\alpha)$, and (iii) $\max\{c_{n,b}(1-\alpha/2), c_\infty(1-\alpha/2) + \kappa^*_{ET}(\alpha)\}$ and $\min\{c_{n,b}(\alpha/2), c_\infty(\alpha/2) - \kappa^*_{ET}(\alpha)\}$, respectively.

By definition, the SC factors $\kappa_{ET,Fix}(\alpha)$ $(\in [0, \infty))$, $\kappa_{ET}(\alpha)$ $(\in [0, \infty))$, and $\kappa^*_{ET}(\alpha)$ $(\in \{-\infty\} \cup [0, \infty))$, respectively, are the smallest values that satisfy

$$\sup_{h \in H}[1 - J_h((c_{Fix}(1-\alpha/2) + \kappa_{ET,Fix}(\alpha))-) + J_h(c_{Fix}(\alpha/2) - \kappa_{ET,Fix}(\alpha))] \leq \alpha,$$

$$\sup_{(g,h)\in GH}[1 - J_h((c_g(1-\alpha/2) + \kappa_{ET}(\alpha))-) + J_h(c_g(\alpha/2) - \kappa_{ET}(\alpha))] \leq \alpha, \text{ and}$$
$$\sup_{(g,h)\in GH}[1 - J_h(\max\{c_g(1-\alpha/2), c_\infty(1-\alpha/2) + \kappa^*_{ET}(\alpha)\}-) +$$
$$J_h(\min\{c_g(\alpha/2), c_\infty(\alpha/2) - \kappa^*_{ET}(\alpha)\})] \leq \alpha. \quad (7.1)$$

(If no such smallest value exists, we take some value that is arbitrarily close to the infimum. If no value that satisfies (7.1) exists, then size-correction is not possible.)

For each test, the condition in (7.1) guarantees that the "overall" asymptotic size of the test is less than or equal to α. It does not guarantee that the maximum (asymptotic) rejection probability in each tail is less than or equal to $\alpha/2$. If the latter is desired, then one should size correct the lower and upper critical values of the equal-tailed test in the same way as one-sided t tests are size corrected in the paper. (This can yield the overall size of the test to be strictly less than α if the (g, h) vector that maximizes the rejection probability is different for the lower and upper critical values.)

Given SC factors that satisfy (7.1), the equal-tailed SC-FCV, SC-Sub, and SC-Hyb t tests have $AsySz(\theta_0) \leq \alpha$ under Assumptions A-E, G, and J by the proofs of Corollary 2 of AG1 and Corollary 1 of the paper. (Only Assumptions A and B are needed for the SC-FCV tests.) Under continuity conditions on $J_h(x)$ at suitable (h, x) such that the inequalities in (7.1) hold as equalities, these tests have $AsySz(\theta_0) = \alpha$.

An alternative way of size-correcting equal-tailed tests is the following method. This method has the advantage that if it is possible to produce an *equal-tailed* size-corrected test, then the procedure does so. Its disadvantage is that it is somewhat more complicated to implement.

First, let $\kappa_{ET,Fix,1}(\alpha) \in [0,\infty)$, $\kappa_{ET,1}(\alpha) \in [0,\infty)$, and $\kappa^*_{ET,1}(\alpha) \in \{-\infty\} \cup [0,\infty)$ denote the smallest values such that

$$\sup_{h\in H}[1 - J_h((c_{Fix}(1-\alpha/2) + \kappa_{ET,Fix,1}(\alpha))-)] \leq \alpha/2,$$
$$\sup_{(g,h)\in GH}(1 - J_h((c_g(1-\alpha/2) + \kappa_{ET,1}(\alpha))-)) \leq \alpha/2, \text{ and}$$
$$\sup_{(g,h)\in GH}\left(1 - J_h\left(\max\{c_g(1-\alpha/2), c_\infty(1-\alpha/2) + \kappa^*_{ET,1}(\alpha)\}-\right)\right) \leq \alpha/2. \quad (7.2)$$

Next, let $\kappa_{ET,Fix,2}(\alpha) \in R$, $\kappa_{ET,2}(\alpha) \in R$, and $\kappa^*_{ET,2}(\alpha) \in \{-\infty\} \cup R$ denote the smallest values such that

$$\sup_{h\in H}[1 - J_h((c_{Fix}(1-\alpha/2) + \kappa_{ET,Fix,1}(\alpha))-) + J_h(c_{Fix}(\alpha/2) - \kappa_{ET,Fix,2}(\alpha))] \leq \alpha,$$
$$\sup_{(g,h)\in GH}[1 - J_h((c_g(1-\alpha/2) + \kappa_{ET,1}(\alpha))-) + J_h(c_g(\alpha/2) - \kappa_{ET,2}(\alpha))] \leq \alpha, \text{ and}$$
$$\sup_{(g,h)\in GH}[1 - J_h(\max\{c_g(1-\alpha/2), c_\infty(1-\alpha/2) + \kappa^*_{ET,1}(\alpha)\}-) +$$
$$J_h(\min\{c_g(\alpha/2), c_\infty(\alpha/2) - \kappa^*_{ET,2}(\alpha)\})] \leq \alpha. \quad (7.3)$$

The "alternative" SC equal-tailed FCV test rejects H_0 if $T_n(\theta_0) > c_{Fix}(1-\alpha/2) + \kappa_{ET,Fix,1}(\alpha)$ or $T_n(\theta_0) < c_{Fix}(\alpha/2) - \kappa_{ET,Fix,2}(\alpha)$. The "alternative" SC equal-tailed Sub and Hyb tests are defined analogously. We use "alternative" SC equal-tailed tests in the Parameter of Interest Near a Boundary Example in Andrews and Guggenberger (2005). For all of the other examples, we use the SC equal-tailed tests defined in (7.1).

If a parameter γ_2 appears in γ and γ_2 is consistently estimable, then PSC tests are more powerful asymptotically than SC tests (because they lead to smaller critical values under some distributions but still have correct asymptotic size). Equal-tailed (i) PSC-FCV, (ii) PSC-Sub, and (iii) PSC-Hyb tests are defined as the SC versions are defined above, but with $\kappa_{ET,Fix}(\alpha)$, $\kappa_{ET}(\alpha)$, and $\kappa^*_{ET}(\alpha)$ replaced by $\kappa_{ET,Fix,\widehat{\gamma}_{n,2}}(\alpha)$, $\kappa_{ET,\widehat{\gamma}_{n,2}}(\alpha)$, and $\kappa^*_{ET,\widehat{\gamma}_{n,2}}(\alpha)$, respectively. Here, the PSC factors $\kappa_{ET,Fix,h_2}(\alpha)$ $(\in [0,\infty))$, $\kappa_{ET,h_2}(\alpha)$ $(\in [0,\infty))$, and $\kappa^*_{ET,h_2}(\alpha)$ $(\in \{-\infty\} \cup [0,\infty))$ are defined to be the smallest values that satisfy:

$$\begin{aligned}
&\sup_{h_1 \in H_1} [1 - J_{(h_1,h_2)}((c_{Fix}(1-\alpha/2) + \kappa_{ET,Fix,h_2}(\alpha))-) + \\
&J_{(h_1,h_2)}(c_{Fix}(\alpha/2) - \kappa_{ET,Fix,h_2}(\alpha))] \le \alpha, \\
&\sup_{g_1,h_1 \in H_1 : ((g_1,h_2),(h_1,h_2)) \in GH} [1 - J_{(h_1,h_2)}((c_{(g_1,h_2)}(1-\alpha/2) + \kappa_{ET,h_2}(\alpha))-) + \\
&J_{(h_1,h_2)}(c_{(g_1,h_2)}(\alpha/2) - \kappa_{ET,h_2}(\alpha))] \le \alpha, \text{ and} \\
&\sup_{g_1,h_1 \in H_1 : ((g_1,h_2),(h_1,h_2)) \in GH} [1 - J_{(h_1,h_2)}(\max\{c_{(g_1,h_2)}(1-\alpha/2), c_\infty(1-\alpha/2) + \\
&\kappa^*_{ET,h_2}(\alpha)\}-) + J_{(h_1,h_2)}(\min\{c_{(g_1,h_2)}(\alpha/2), c_\infty(\alpha/2) - \kappa^*_{ET,h_2}(\alpha)\})] \le \alpha. \quad (7.4)
\end{aligned}$$

The (i) PSC-FCV, (ii) PSC-Sub, and (iii) PSC-Hyb equal-tailed tests all have $AsySz(\theta_0) \le \alpha$ under Assumptions N plus (i) A and B, (ii) A-E, G, and J, and (iii) A-E, G, J, and K, respectively. (The proof is analogous to the proof of Theorem 3 of the paper combined with the proof of Theorem 2 of the paper.) These tests have $AsySz(\theta_0) = \alpha$ provided the inequalities in (7.4) hold as equalities.

The finite-sample adjustments introduced in Section 4 of the paper do not cover equal-tailed tests. For equal-tailed subsampling tests, we define the following finite-sample adjustment to $AsySz(\theta_0)$:

$$AsySz_n(\theta_0) = \sup_{h \in H} [1 - J_h(c_{(\delta_n^r h_1, h_2)}(1-\alpha/2)-) + J_h(c_{(\delta_n^r h_1, h_2)}(\alpha/2))]. \quad (7.5)$$

Define $Max^{r-}_{ET,Sub}(\alpha)$ as $Max_{ET,Hyb}(\alpha)$ is defined in (5.2) of the paper with $c^*_g(1-\alpha/2)$ and $c^{**}_g(\alpha/2)$ replaced by $c_g(1-\alpha/2)$ and $c_g(\alpha/2)-$, respectively, where " $-$ " indicates the limit from the left. Define $Max^{\ell-}_{ET,Sub}(\alpha)$ analogously with $c^*_g(1-\alpha/2)$ and $c^{**}_g(\alpha/2)$ replaced by $c_g(1-\alpha/2)-$ and $c_g(\alpha/2)$. With the function that appears in Assumption P(i) altered to $(g,h) \to J_h(c_g(1-\alpha/2)-) - J_h(c_g(\alpha/2))$ and with $Max^{r-}_{ET,Sub}(\alpha) = Max^{\ell-}_{ET,Sub}(\alpha)$ in place of Assumption P(ii), the result of Theorem 4(a) of the paper, viz., $AsySz_n(\theta_0) \to AsySz(\theta_0)$, holds for equal-tailed subsampling tests. An analogous result holds for equal-tailed hybrid tests.

Based on (7.5), we introduce finite-sample adjustments that can improve the asymptotic approximations upon which the equal-tailed SC and PSC subsampling and hybrid tests rely. Equal-tailed ASC and APSC subsampling and hybrid tests are defined just as SC and PSC subsampling and hybrid tests are defined, but using $\kappa_{ET}(\delta_n, \alpha)$, $\kappa^*_{ET}(\delta_n, \alpha)$, $\kappa_{ET,\widehat{\gamma}_{n,2}}(\delta_n, \alpha)$, and $\kappa^*_{ET,\widehat{\gamma}_{n,2}}(\delta_n, \alpha)$ in place of $\kappa_{ET}(\alpha)$ and $\kappa^*_{ET}(\alpha)$. The ASC factors $\kappa_{ET}(\delta, \alpha)$ ($\in [0, \infty)$) and $\kappa^*_{ET}(\delta, \alpha)$ ($\in \{-\infty\} \cup [0, \infty)$) are defined to be the smallest values that satisfy:

$$\begin{aligned}
&\sup_{(h_1,h_2)\in H} [1 - J_{(h_1,h_2)}((c_{(\delta^r h_1,h_2)}(1-\alpha/2) + \kappa_{ET}(\delta,\alpha))-) + \\
&J_{(h_1,h_2)}(c_{(\delta^r h_1,h_2)}(\alpha/2) - \kappa_{ET}(\delta,\alpha))] \leq \alpha \text{ and} \\
&\sup_{(h_1,h_2)\in H} [1 - J_{(h_1,h_2)}(\max\{c_{(\delta^r h_1,h_2)}(1-\alpha/2), c_\infty(1-\alpha/2) + \kappa^*_{ET}(\delta,\alpha)\}-) + \\
&J_{(h_1,h_2)}(\min\{c_{(\delta^r h_1,h_2)}(\alpha/2), c_\infty(\alpha/2) - \kappa^*_{ET}(\delta,\alpha)\})] \leq \alpha. \qquad (7.6)
\end{aligned}$$

The APSC factors $\kappa_{ET,h_2}(\delta, \alpha)$ ($\in [0, \infty)$) and $\kappa^*_{ET,h_2}(\delta, \alpha)$ ($\in \{-\infty\} \cup [0, \infty)$) are defined to be the smallest values that satisfy:

$$\begin{aligned}
&\sup_{h_1\in H_1} [1 - J_{(h_1,h_2)}((c_{(\delta^r h_1,h_2)}(1-\alpha/2) + \kappa_{ET,h_2}(\delta,\alpha))-) + \\
&J_{(h_1,h_2)}(c_{(\delta^r h_1,h_2)}(\alpha/2) - \kappa_{ET,h_2}(\delta,\alpha))] \leq \alpha \text{ and} \\
&\sup_{h_1\in H_1} [1 - J_{(h_1,h_2)}(\max\{c_{(\delta^r h_1,h_2)}(1-\alpha/2), c_\infty(1-\alpha/2) + \kappa^*_{ET,h_2}(\delta,\alpha)\}-) + \\
&J_{(h_1,h_2)}(\min\{c_{(\delta^r h_1,h_2)}(\alpha/2), c_\infty(\alpha/2) - \kappa^*_{ET,h_2}(\delta,\alpha)\})] \leq \alpha. \qquad (7.7)
\end{aligned}$$

The ASC and APSC tests have $AsySz(\theta_0) = \alpha$ under conditions that are similar to those given in Section 4 of the paper. For brevity, we do not give details.

8 Size-Corrected Combined Test

Theorem S1(c)(iv)-(vi) and Figure B-2(f) show that in some contexts the SC-Hyb test can be more powerful than the SC-Sub test for some $(g, h) \in GH$ and vice versa for other $(g, h) \in GH$. This implies that a test that combines the SC-Hyb and SC-Sub tests can be more powerful than both. In this Section, we introduce such a test. It is called the size-corrected combined (SC-Com) test. This test has power advantages over the SC-Hyb and SC-Sub tests in some cases. This is illustrated in Figure B-2(f) where the critical value function of the SC-Com test is the minimum of the upper horizontal SC-Hyb critical value function and the upper curved SC-Sub critical value function. On the other hand, the SC-Com test has computational disadvantages because it requires computation of the critical values for both the SC-Sub and SC-Hyb tests, which requires calculation of $\kappa(\alpha)$ and $\kappa^*(\alpha)$ in cases where both the subsampling and hybrid tests need size correction. Furthermore, in most contexts, the SC-Hyb test is more powerful than the SC-Sub for all $(g,h) \in GH$, so the SC-Com test just reduces to the SC-Hyb test.

The size-corrected combined (SC-Com) test rejects $H_0 : \theta = \theta_0$ when

$$T_n(\theta_0) > c_{n,Com}(1-\alpha), \text{ where} \tag{8.1}$$
$$c_{n,Com}(1-\alpha) = \min\{c_{n,b}(1-\alpha) + \kappa(\alpha), \max\{c_{n,b}(1-\alpha), c_\infty(1-\alpha) + \kappa^*(\alpha)\}\},$$

where the constants $\kappa(\alpha)$ and $\kappa^*(\alpha)$ are defined in (3.2) of the paper.

The following result shows that the SC-Com test has $AsySz(\theta_0) = \alpha$.

Theorem S2. *Suppose Assumptions* A-G *and* K-M *hold. Then, the* SC-Com *test satisfies* $AsySz(\theta_0) = \alpha$.

Comments. 1. By definition, the critical value, $c_{n,Com}(1-\alpha)$, of the SC-Com test is less than or equal to those of the SC-Sub and SC-Hyb tests. By (3.2) of the paper, it is less than or equal to that of the SC-FCV test as well. Hence, the SC-Com test is at least as powerful as the SC-Sub, SC-Hyb, and SC-FCV tests.

2. A PSC-Com test can be defined as in (8.1) with $\kappa(\alpha)$ and $\kappa^*(\alpha)$ replaced by $\kappa_{\widehat{\gamma}_{n,2}}(\alpha)$, and $\kappa^*_{\widehat{\gamma}_{n,2}}(\alpha)$.

3. An ASC-Com test can be defined as in (8.1) with $\kappa(\alpha)$ and $\kappa^*(\alpha)$ replaced by $\kappa(\delta_n, \alpha)$ and $\kappa^*(\delta_n, \alpha)$, respectively. Suppose Assumptions A-G, K-M, and Q hold. Then, the ASC-Com test satisfies $AsySz(\theta_0) = \alpha$. This holds by the argument in the proof of Theorem 4(b) of the paper (see above) using the results of Theorem 4(b)(i) of the paper.

4. An APSC-Com test can be defined as in (8.1) with $\kappa(\alpha)$ and $\kappa^*(\alpha)$ replaced by $\kappa_{\widehat{\gamma}_{2,n}}(\delta_n, \alpha)$ and $\kappa^*_{\widehat{\gamma}_{2,n}}(\delta_n, \alpha)$, respectively. Suppose Assumptions A-G, K, L, N, O, R, and Q hold. Then, the APSC-Com test satisfies $AsySz(\theta_0) = \alpha$. This holds by the argument in the proof of Theorem 4(c) of the paper (see above) using the results of Theorem 4(c)(i) of the paper.

Proof of Theorem S2. By the same argument as in the proof of Theorem 1(ii) of AG1, the SC-Com test satisfies

$$AsySz(\theta_0) \tag{8.2}$$
$$\leq \sup_{(g,h)\in GH} [1 - J_h(\min\{c_g(1-\alpha) + \kappa(\alpha), \max\{c_g(1-\alpha), c_\infty(1-\alpha) + \kappa^*(\alpha)\}\}-)].$$

By the proof of Theorem 2 of the paper, the constants $\kappa(\alpha)$ and $\kappa^*(\alpha)$ defined in (3.2) of the paper are such that (2.7) above holds and hence for all $(g,h) \in GH$,

$$1 - J_h(c_g(1-\alpha) + \kappa(\alpha)-) \leq \alpha \text{ and}$$
$$1 - J_h(\max\{c_g(1-\alpha), c_\infty(1-\alpha) + \kappa^*(\alpha)\}-) \leq \alpha. \tag{8.3}$$

Equations (8.2) and (8.3) combine to give $AsySz(\theta_0) \leq \alpha$.

The SC-Com test has $AsySz(\theta_0) \geq \alpha$ because its $AsySz(\theta_0)$ is greater than or equal to that of the SC-Sub test (because its critical value is no larger) and the latter equals α by Theorem 2 of the paper. □

9 Testing When a Nuisance Parameter May Be Near a Boundary

This example is a continuation of an example in AG1. It is a testing problem where a nuisance parameter may be near the boundary of the parameter space under the null hypothesis. The observations are $\{X_i \in R^2 : i \leq n\}$, which are i.i.d. with distribution F, $X_i = (X_{i1}, X_{i2})'$, $E_F X_i = (\theta, \mu)'$, and (X_{i1}, X_{i2}) have correlation ρ. The null hypothesis is $H_0 : \theta = 0$, i.e., $\theta_0 = 0$. The parameter space for the nuisance parameter μ is $[0, \infty)$. The test statistic $T_n(\theta_0)$ equals $T_n^*(\theta_0)$, $-T_n^*(\theta_0)$, or $|T_n^*(\theta_0)|$, where $T_n^*(\theta_0)$ is a t statistic based on the Gaussian quasi-ML estimator of θ that imposes the restriction that $\mu \in [0, \infty)$ and uses any consistent estimators of the standard deviations and correlation of X_{i1} and X_{i2} in the quasi-likelihood, see AG1 for details. In AG1, Assumptions A-G are verified.

Table B-I reports maximum (over $h_1 = \lim_{n\to\infty} n^{1/2}\mu_{n,h}/\sigma_{n,h,2}$) null rejection probabilities (as percentages) for several fixed values of h_2 $(= \lim_{n\to\infty} \rho_{n,h})$ for hybrid and several other nominal 5% tests.[2] Depending on the column, the probabilities are asymptotic or finite-sample. The finite-sample results are for the case of $n = 120$ and $b = 12$ with $\widehat{\sigma}_{n1}, \widehat{\sigma}_{n2}$, and $\widehat{\rho}_n$ being the sample standard deviations and correlation of X_{i1} and X_{i2}. To dramatically increase computational speed, here and in all of the tables below, finite-sample subsampling and hybrid results are based on $q_n = 119$ subsamples of consecutive observations.[3] Hence, only a small fraction of the "120 choose 12" available subsamples are used. In cases where subsampling and hybrid tests have correct asymptotic size, their finite-sample performance is expected to be better when all available subsamples are used than when only $q_n = 119$ are used. This should be taken into account when assessing the results of the tables. Panels (a), (b), and (c) of Table B-I give results for upper one-sided, symmetric two-sided, and equal-tailed two-sided tests, respectively. The results for lower one-sided tests are the same as for the upper tests with the sign of h_2 changed (by symmetry) and, hence, are not given. The rows labelled Max give the size (asymptotic or $n = 120$) of the test considered. For brevity, we refer below to the numbers given in the tables

[2] The finite sample results in Table B-I are based on $20,000$ simulation repetitions. The asymptotic results are based on $50,000$ simulation repetitions. For the asymptotic results, the search over h_1 is done with stepsize 0.05 on $[0, 10]$ and also includes the value $h_1 = 9,999,999,999$. For the finite-sample results, the search over h_1 is done with stepsize .001 on $[0, 0.5]$, with stepsize 0.05 on $[0.5, 1.0]$, and with stepsize 1.0 on $[1.0, 10]$. Calculations indicate that these stepsizes are sufficiently small to yield accuracy to within $\pm.1$.

[3] This includes 10 "wrap-around" subsamples that contain observations at the end and beginning of the sample, for example, observations indexed by $(110, ..., 120, 1)$. The choice of $q_n = 119$ subsamples is made because this reduces rounding errors when q_n is small when computing the sample quantiles of the subsample statistics. The values ν_α that solve $\nu_\alpha/(q_n + 1) = \alpha$ for $\alpha = .025$, .95, and .975 are the integers 3, 114, and 117. In consequence, the .025, .95, and .975 sample quantiles are given by the 3rd, 114th, and 117th largest subsample statistics. See Hall (1992, p. 307) for a discussion of this choice in the context of the bootstrap.

as though they are precise, but of course they are subject to simulation error.

9.1 FCV, Subsampling, and Hybrid Tests

Column 2 of Table B-I shows that subsampling tests have very large asymptotic size distortions for upper one-sided and equal-tailed two-sided tests (nominal 5% tests have asymptotic levels 50.2 and 52.7%, respectively), and moderate size distortions for symmetric two-sided tests (the nominal 5% test has asymptotic level 10.1%). Also, column 7 of Table B-I shows that the FCV tests have very small asymptotic size distortions for upper one-sided tests (the nominal 5% test has asymptotic level 5.8%), and no size distortions for symmetric and equal-tailed two-sided tests.

Column 10 of Table B-I shows that the nominal 5% hybrid test has asymptotic size of 5% for upper, symmetric, and equal-tailed tests. So, the hybrid test has correct asymptotic size for all three types of tests in this example.

Finite-sample results for the Sub, FCV, and Hyb tests are given in columns 4, 8, and 12 of Table B-I, respectively. For Hyb tests, the asymptotic approximations are fairly accurate, but tend to over-estimate the finite-sample rejection rates somewhat for some values of h_2 with finite-sample values varying between 3.4 and 5.2% compared to the asymptotic values of 5.0%. For FCV tests, the asymptotic approximations are found to be very accurate for upper tests and quite accurate for symmetric and equal-tailed tests.

The asymptotic approximations for the Sub test are found to be quite good for h_2 values where the (maximum) asymptotic probabilities equal 5.0%. But, for h_2 values where they exceed 5.0%, they tend to over-estimate the finite-sample values—sometimes significantly so, e.g., 33.8 versus 25.6% for $h_2 = -.95$ with upper Sub tests. Nevertheless, in the worst case scenarios (i.e., for h_2 values of 1.0 or -1.0, which yield the greatest asymptotic rejection probabilities), the asymptotic approximations are quite good. Hence, the asymptotic sizes and finite-sample sizes are close—50.2 versus 49.8%, 10.1 versus 8.4%, and 52.7 versus 52.7% for upper, symmetric, and equal-tailed tests, respectively.

The results in Table B-I for the columns headed Adj-Asy, PSC-Sub, APSC-Sub, ... are discussed below.

9.2 Plug-in Tests

The upper, symmetric, and equal-tailed subsampling tests and the upper FCV test need size-correction in this example. Plug-in size correction is possible because estimation of the correlation parameter ρ is straightforward using the usual sample correlation estimator. Columns 5 and 9 of Table B-I provide the finite-sample (maximum) rejection probabilities of the nominal 5% PSC-Sub and PSC-FCV tests. Results for the symmetric and equal-tailed PSC-FCV tests are not given because the PSC-FCV and FCV tests are the same in these cases since the FCV test has correct asymptotic size. Results for the PSC-Hyb test are not given because it is the same as the Hyb test.

The results for the PSC-Sub tests are impressive. The finite-sample sizes of the upper, symmetric, and equal-tailed tests are 5.3, 5.1, and 5.5%, respectively, whereas the finite-sample sizes of the Sub tests are 49.8, 8.4, and 52.7%. The plug-in feature of the size-correction method yields (maximum) rejection probabilities across different h_2 values that are all reasonably close to 5.0%—ranging from 3.1 to 5.5%, with most being between 4.5 and 5.5%. Having these values all close to 5% is desirable from a power perspective.

The upper FCV test only requires minor size-correction given that its asymptotic and finite-sample size is 5.8%. The PSC-FCV test provides improvement. Its finite-sample size is 5.2%.

9.3 Finite-Sample Adjusted Tests

Column 3 of Table B-I gives the finite-sample adjusted asymptotic rejection probabilities ($\times 100$) of the subsampling test. These values are noticeably closer to the finite-sample values given in column 4 than are the (unadjusted) asymptotic rejection probabilities given in column 2. For example, for the upper subsampling test and $h_2 = -.95$, the values for Adj-Asy, $n = 120$, and Asy are 22.9, 25.6, and 33.8%, respectively. Hence, the adjustment works pretty well for the subsampling test here. For the hybrid test, the adjusted asymptotic and unadjusted asymptotic rejection rates are all 5.0%. So, the adjustment makes no difference for the hybrid test in this example.

Column 6 of Table B-I reports the finite-sample rejection probabilities of the APSC-Sub test. For upper and equal-tailed tests, the adjustment leads to over-correction of the Sub test when the finite-sample correlation, denoted here by h_2, is close to -1 and 1, respectively, and appropriate size-correction for other values of h_2. In consequence, for these cases the PSC-Sub test (see column 5) has better finite-sample size (viz., 5.3 and 5.5%) than the APSC-Sub test (13.5 and 13.5%). For symmetric tests, both of these size-corrected tests perform well.

In conclusion, in this example, the hybrid and PSC-Sub tests perform quite well in terms of finite-sample size for upper, symmetric, and equal-tailed tests. The APSC-Sub test performs well for symmetric test, but not so well for upper and equal-tailed tests.

10 Autoregression Example

This section provides results for the conditionally heteroskedastic autoregression example.

10.1 Upper and Lower One-sided CIs

First, we discuss Table B-II, which is analogous to Table II of the paper but provides results for upper and lower one-sided CIs rather than symmetric and equal-tailed two-sided CIs. (See the footnote to Table II of the paper and Section 1 above for details concerning the construction of Table II, which also apply to Table B-II.)

Due to the asymmetry of the asymptotic distribution J_h^* of the test statistic $T_n^*(\theta_0)$, the results for upper and lower one-sided CIs are quite different. Table B-II shows that upper one-sided FCV CIs have correct asymptotic size (up to simulation error). The same is true of upper one-sided hybrid CIs. Upper one-sided subsampling CIs, however, exhibit substantial asymptotic size distortion. The Adj-Asy size distortion of the subsampling CIs is noticeably smaller than the asymptotic size distortion and the former gives a better approximation to the finite-sample size distortion for sample size $n = 131$. The reason for the results just described for the upper one-sided FCV, hybrid, and subsampling CIs is that the upper tail of the asymptotic distribution J_h^* gets thinner as h_1 goes to zero. In consequence, the $1 - \alpha$ quantile of J_h^* is increasing in h_1, which leads to size distortion for the subsampling CI but not the FCV CI.

For lower one-sided CIs, the opposite is true. The lower tail of J_h^* gets thicker as h_1 goes to zero. In consequence, the lower one-sided FCV exhibits substantial asymptotic size distortion, whereas the subsampling and hybrid CIs have correct asymptotic size.

10.2 Verification of Assumptions for CI for an Autoregressive Parameter

In this section, we verify the assumptions of Corollary 2 of the paper, viz., Assumptions A-G, J, K-M, T, and TET, for the AR(1) Example. We use Lemma 4 of AG1 to verify Assumption G. Lemma 4 of AG1 requires verification of Assumptions t1, Sub1, A, BB, C, DD, EE, and HH. Note that the latter assumptions imply Assumptions B and D. Corollary 2 of the paper establishes the desired results for the hybrid test. For the FCV and subsampling tests, the desired results hold under the same conditions by Corollary 1 in Appendix A2 of Andrews and Guggenberger (2009b).

Assumptions BB(a) and (c) are verified by Proposition S1 stated below that is proved in Andrews and Guggenberger (2008), hereafter AG-AR. Assumptions t1, Sub1, A, C, DD, F, J, T, M, TET, K, L, and BB(b) are verified in the next subsection. Verifications of Assumptions E and EE are given in Sections 10.2.4 and

10.2.5 below for model 1. For brevity, we do not verify these assumptions for model 2. Finally, Assumption HH is verified in Section 10.2.6.

10.2.1 Verification of Assumptions t1, Sub1, A, C, DD, F, J, T, M, TET, K, L, and BB(b)

Assumption t1 holds with $\tau_n = n^{1/2}$ by definition of $T_n^*(\theta_0)$. Assumptions Sub1 and A clearly hold. Assumption C holds by the choice of b_n. Assumption DD holds when the AR parameter is less than one by the assumption of a strictly stationary initial condition. In the unit root case, it holds by the i.i.d. assumption on the innovations for $i = 1, ..., n$ and the fact that the test statistic $T_n^*(\theta_0)$ is invariant to the initial condition. Assumption F holds because J_h^* and $-J_h^*$ are strictly increasing on R for all $h \in H$ and $|J_h^*|$ is strictly increasing on R_+ and has support R_+ for all $h \in H$. For the same reason, Assumption J holds for $J_h = J_h^*$. Assumption T holds for $J_h = J_h^*$ and $J_h = -J_h^*$ because J_h^* is continuous on R and has support R for all $h \in H$. Assumption T holds for $J_h = |J_h^*|$ because $|J_h^*|$ is continuous on R_+ and has support R_+ for all $h \in H$. For the same reasons, Assumptions M(a)(ii), M(b)(ii), and M(c)(ii) hold for $J_h = J_h^*, -J_h^*$, and $|J_h^*|$ and Assumption TET holds.

Assumption K holds because J_h^* is $N(0, 1)$ for all $h = (h_1, h_2) \in H$ with $h_1 = \infty$.

Assumption L holds by properties of the Ornstein-Uhlenbeck process. Numerical calculations indicate that the supremum and infimum in this assumption are attained at $h_1 = 0$ or $h_1 = \infty$ (depending upon whether the supremum or infimum is being considered and whether $J_h = J_h^*$, $-J_h^*$, or $|J_h^*|$). This indicates that Assumption M(a)(i) holds. Numerical calculations also indicate that the supremum in Assumption M(b)(i) is attained at $h_1 = (0, \infty)$ or $(0, 0)$ for all $h_2 \in H_2$ depending upon whether $J_h = J_h^*$, $-J_h^*$, or $|J_h^*|$ and hence this assumption holds. Assumption M(c)(i) holds because $c_h(1 - \alpha)$ is monotone in h_1 for each $h_2 \in H_2 = \Gamma_2$ (based on numerical calculations), which implies that either H^* is empty or $H^* = \{h \in H : h_1 > 0\}$ depending on whether $J_h = J_h^*, -J_h^*$, or $|J_h^*|$. When H^* is non-empty, $\sup_{h \in H^*} c_h(1-\alpha)$ is attained at $h_1 = \infty$.

Assumption BB(b) holds because $P_\gamma(\widehat{\sigma}_{n,b_n,j} > 0) = 1$ for all $n, b_n \geq 4$, $j = 1, ..., q_n$, and $\gamma \in \Gamma$.

10.2.2 Normalization Constants

In this sub-section, we specify the normalization constants a_n and d_n for which $a_n(\widehat{\rho}_n - \rho_n)$ and $d_n \widehat{\sigma}_n$ have non-degenerate asymptotic distributions under $\{\gamma_{n,h} : n \geq 1\}$. These constants appear in Assumptions BB, EE, and HH. The constants are rather complicated when the innovations exhibit conditional heteroskedasticity. So, we show below how they simplify under conditional homoskedasticity, which should make them easier to interpret.

The normalization constants a_n and d_n depend on $\gamma_{n,h}$ and are denoted $a_n(\gamma_{n,h})$ and $d_n(\gamma_{n,h})$. They are defined as follows. Let $\{w_n : n \geq 1\}$ be any subsequence of

$\{n\}$. Let $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\}$ be a sequence for which $w_n\gamma_{n,1} \to \infty$ or $w_n\gamma_{n,1} \to h_1 < \infty$. Let $\rho_n = 1 - \gamma_{n,1}$. Define the 2-vectors

$$\begin{aligned} X^1 &= (Y^*_{n,0}/\phi_{n,1}, \phi^{-1}_{n,1})' \text{ and} \\ Z &= (1, -E_{F_n}(Y^*_{n,0}/\phi^2_{n,1})/E_{F_n}(\phi^{-2}_{n,1}))'. \end{aligned} \tag{10.1}$$

Define

$$\begin{aligned} a_{w_n}(\gamma_n) &= w_n^{1/2} d_{w_n}(\gamma_n) \text{ and} \\ d_{w_n}(\gamma_n) &= \begin{cases} \frac{E_{F_n}(Y^{*2}_{n,0}/\phi^2_{n,1}) - (E_{F_n}(Y^*_{n,0}/\phi^2_{n,1}))^2/E_{F_n}(\phi^{-2}_{n,1})}{(Z'E_{F_n}(X^1X^{1\prime}U^2_{n,1}/\phi^2_{n,1})Z)^{1/2}} & \text{if } w_n\gamma_{n,1} \to \infty \\ w_n^{1/2} & \text{if } w_n\gamma_{n,1} \to h_1 < \infty. \end{cases} \end{aligned} \tag{10.2}$$

To simplify notation in this paragraph, we delete the subscript n in most expressions below and we omit the subscript F_n on expectations. In the case where $w_n\gamma_{n,1} \to \infty$ and $\rho \to 1$, the constants a_{w_n} and d_{w_n} in (10.2) simplify to

$$a_{w_n} = w_n^{1/2} \frac{E(Y_0^{*2}/\phi_1^2)}{(E(Y_0^{*2}U_1^2/\phi_1^4))^{1/2}} \text{ and } d_{w_n} = \frac{E(Y_0^{*2}/\phi_1^2)}{(E(Y_0^{*2}U_1^2/\phi_1^4))^{1/2}} \tag{10.3}$$

up to lower order terms. This holds because by Lemma S1 below

$$\begin{aligned} Z'E(X^1X^{1\prime}U_1^2/\phi_1^2)Z &= E(Y_0^{*2}U_1^2/\phi_1^4) - 2E(Y_0^*U_1^2/\phi_1^4)E(Y_0^*/\phi_1^2)/E(\phi_1^{-2}) \\ &\quad +(E(Y_0^*/\phi_1^2))^2E(U_1^2/\phi_1^4)/(E(\phi_1^{-2}))^2 \\ &= E(Y_0^{*2}U_1^2/\phi_1^4)(1 + O(1-\rho)) \end{aligned} \tag{10.4}$$

and

$$E(Y_0^{*2}/\phi_1^2) - (E(Y_0^*/\phi_1^2))^2/E(\phi_1^{-2}) = E(Y_0^{*2}/\phi_1^2)(1 + O(1-\rho)). \tag{10.5}$$

If, in addition, $\{U_i : i = ..., 0, 1, ...\}$ are i.i.d. with mean 0, variance $\sigma_U^2 \in (0, \infty)$, and distribution F and $\phi_i = 1$, then the constants a_{w_n} and d_{w_n} simplify to

$$a_{w_n} = w_n^{1/2}(1-\rho_n^2)^{-1/2} \text{ and } d_{w_n} = (1-\rho_n^2)^{-1/2}. \tag{10.6}$$

This follows because in the present case $\phi_i^2 = 1$, $EY_0^{*2} = \sum_{j=0}^{\infty} \rho^{2j} EU_{-j}^2 = (1 - \rho^2)^{-1}\sigma_U^2$, and $E(Y_0^{*2}U_1^2/\phi_1^2) = (1-\rho^2)^{-1}\sigma_U^4$.

Given the definitions of $a_n(\cdot)$ and $d_n(\cdot)$, $\tau_n = a_n(\gamma_{n,h})/d_n(\gamma_{n,h}) = n^{1/2}$ does not depend on $\gamma_{n,h}$, as is required.

10.2.3 Preliminary Results from AG-AR

In this sub-section, we state the result proved in AG-AR that verifies Assumption B of the paper and Assumption BB(a) of AG1. We also state some other results proved in AG-AR because they are used below when verifying Assumptions E, EE, and HH.

We start by stating an assumption, Assumption INNOV, that specifies certain properties for the innovations $U_i = U_{n,i}$ and $\phi_i^2 = \phi_{n,i}^2$. Assumption INNOV automatically holds when one is considering any sequence $\{\gamma_{n,h} : n \geq 1\}$. This follows from the definition of the parameter space $\mathcal{F}(\gamma_2)$ and the definition of a sequence $\{\gamma_{n,h} : n \geq 1\}$. Hence, when showing below that a property holds under a sequence $\{\gamma_{n,h} : n \geq 1\}$, it is sufficient to show that it holds under Assumption INNOV.

Assumption INNOV. (i) For each $n \geq 1$, $\{(U_{n,i}, \phi_{n,i}^2) : i = ..., 0, 1, ...\}$ are stationary and strong mixing with $E(U_{n,i}|\mathcal{G}_{n,i-1}) = 0$ a.s., $E(U_{n,i}^2|\mathcal{G}_{n,i-1}) = \sigma_{n,i}^2$ a.s. where $\mathcal{G}_{n,i}$ is some non-decreasing sequence of σ-fields for $i = ..., 1, 2, ...$ for $n \geq 1$ for which $(U_{n,j}, \phi_{n,j+1}^2) \in \mathcal{G}_{n,i}$ for all $j \leq i$, (ii) the strong-mixing numbers $\{\alpha_n(m) : m \geq 1\}$ satisfy $\alpha(m) = \sup_{n \geq 1} \alpha_n(m) = O(m^{-3\zeta/(\zeta-3)})$ as $m \to \infty$ for some $\zeta > 3$, (iv) $\sup_{n,i,s,t,u,v,A} E_{F_n}|\prod_{a \in A} a|^\zeta < \infty$, where $0 \leq i, s, t, u, v < \infty$, $n \geq 1$, and A is any non-empty subset of $\{U_{n,i-s}, U_{n,i-t}, U_{n,i+1}^2, U_{n,-u}, U_{n,-v}, U_{n,1}^2\}$, (v) $\phi_i^2 \geq \delta > 0$ *a.s.*, (vi) $\lambda_{\min} E(X^1 X^{1\prime} U_{n,1}^2/\phi_{n,1}^2) \geq \delta > 0$, where $X^1 = (Y_{n,0}^*/\phi_{n,1}, \phi_{n,1}^{-1})'$, and (vii) the following limits exist and are positive: $h_{2,1} = \lim_{n\to\infty} EU_{n,i}^2$, $h_{2,2} = \lim_{n\to\infty} E(U_{n,i}^2/\phi_{n,i}^4)$, $h_{2,3} = \lim_{n\to\infty} E(U_{n,i}^2/\phi_{n,i}^2)$, $h_{2,4} = \lim_{n\to\infty} E\phi_{n,i}^{-1}$, $h_{2,5} = \lim_{n\to\infty} E\phi_{n,i}^{-2}$, and $h_{2,6} = \lim_{n\to\infty} E\phi_{n,i}^{-4}$.

Given that $\phi_{n,i}$ is bounded away from zero by Assumption INNOV(v), Assumption INNOV(iv) implies that $\sup_{n,i,s,t,u,v,A^*} E_{F_n}|\prod_{a\in A^*} a|^\zeta < \infty$, where $0 \leq i, s, t, u, v < \infty$, $n \geq 1$, and A^* is a non-empty subset of $\{U_{n,i-s}, U_{n,i-t}, U_{n,i+1}^2/\phi_{n,i+1}^4, U_{n,-u}, U_{n,-v}, U_{n,1}^2/\phi_{n,1}^4\}$ or a subset of $\{U_{n,i-s}, U_{n,i-t}, \phi_{n,i+1}^{-k}, U_{n,-u}, U_{n,-v}, \phi_{n,1}^{-k}\}$ for $k = 2, 3, 4$. The uniform bound on these expectations is needed in the application of the mixing inequality in (10.15) used below in the verification of Assumption E.

Define $h_{n,1}$ by $\gamma_{n,h,1} = h_{n,1}/n$. Then, $h_{n,1} \to h_1$ as $n \to \infty$ because $n\gamma_{n,h,1} \to h_1$. In this example, $h_{n,1} = 0$ corresponds to a unit root, i.e., $\rho_n = 1 - \gamma_{n,h,1} = 1 - h_{n,1}/n = 1$. If $h_{n,1} = 0$, then the initial condition $Y_{n,0}^*$ is arbitrary. If $h_{n,1} > 0$, then under the assumptions in the paper the initial condition satisfies the following stationarity condition:

Assumption STAT. $Y_{n,0}^* = \sum_{j=0}^\infty \rho_n^j U_{n,-j}$, where $\rho_n = 1 - h_{n,1}/n$.

Let $W(\cdot)$ and $W_2(\cdot)$ be independent standard Brownian motions on $[0, 1]$ and let Z_1 be a standard normal random variable that is independent of $W(\cdot)$ and $W_2(\cdot)$. By definition,

$$I_h(r) = \int_0^r \exp(-(r-s)h_1)dW(s),$$

$$I_h^*(r) = I_h(r) + \frac{1}{\sqrt{2h_1}} \exp(-h_1 r)Z_1 \text{ for } h_1 > 0 \text{ and } I_h^*(r) = W(r) \text{ for } h_1 = 0,$$

$$I^*_{D,h}(r) = I^*_h(r) - \int_0^1 I^*_h(s)ds, \text{ and}$$

$$Z_2 = \left(\int_0^1 I^*_{D,h}(r)^2 dr \right)^{-1/2} \int_0^1 I^*_{D,h}(r) dW_2(r). \tag{10.7}$$

Note that Z_2 has a $N(0,1)$ distribution conditional on $(Z_1, W(\cdot))$. Hence, Z_2 has an unconditional $N(0,1)$ distribution and is independent of $(Z_1, W(\cdot))$.

AG-AR prove the following Proposition.

Proposition S1. *Suppose* (i) *Assumption* INNOV *holds,* (ii) *Assumption* STAT *holds when* $\rho_n < 1$, (iii) $\rho_n \in [-1+\varepsilon, 1]$ *for some* $0 < \varepsilon < 2$, *and* (iv) $\rho_n = 1 - h_{n,1}/n$ *and* $h_{n,1} \to h_1 \in [0, \infty]$. *Then,*

$$a_n(\widehat{\rho}_n - \rho_n) \to_d V_h, \; d_n \widehat{\sigma}_n \to_d Q_h, \text{ and } \frac{a_n(\widehat{\rho}_n - \rho_n)}{d_n \widehat{\sigma}_n} \to_d J_h,$$

where a_n, d_n, V_h, Q_h, *and* J_h *are defined as follows.*
(a) *In model* 1, *for* $h_1 \in [0, \infty)$, $a_n = n$, $d_n = n^{1/2}$, V_h *is the distribution of*

$$h_{2,7} \frac{\int_0^1 I^*_{D,h}(r) dW(r)}{h_{2,2}^{1/2} h_{2,1}^{1/2} \int_0^1 I^*_{D,h}(r)^2 dr} + (1 - h_{2,7}^2)^{1/2} \frac{\int_0^1 I^*_{D,h}(r) dW_2(r)}{h_{2,2}^{1/2} h_{2,1}^{1/2} \int_0^1 I^*_{D,h}(r)^2 dr}, \tag{10.8}$$

Q_h *is the distribution of*

$$h_{2,2}^{-1/2} h_{2,1}^{-1/2} \left[\int_0^1 I^*_{D,h}(r)^2 dr \right]^{-1/2}, \tag{10.9}$$

and J_h *is the distribution of*

$$h_{2,7} \frac{\int_0^1 I^*_{D,h}(r) dW(r)}{\left(\int_0^1 I^*_{D,h}(r)^2 dr \right)^{1/2}} + (1 - h_{2,7}^2)^{1/2} Z_2. \tag{10.10}$$

(b) *In model* 1, *for* $h_1 = \infty$, a_n *and* d_n *are defined as in* (10.2) *with* n *in place of* w_n, V_h *is a* $N(0,1)$ *distribution,* Q_h *is the distribution of the constant one, and* J_h *is a* $N(0,1)$ *distribution.*

In the remainder of this sub-section, we state several other results that are proved in AG-AR and are used below when verifying Assumptions E, EE, and HH. In the proof of Proposition S1 for the case $n(1-\rho) \to \infty$, AG-AR show the following results. If $\rho \to 1$,

$$\frac{n^{-1/2} X_1' P_{X_2} U}{(E(Y_0^{*2} U_1^2 / \phi_1^4))^{1/2}} \to_p 0 \text{ and } \sum_{i=1}^n \zeta_i \to_d N(0,1), \text{ where}$$

$$\zeta_i = n^{-1/2} \frac{Y^*_{i-1} U_i / \phi_i^2}{(E(Y_0^{*2} U_1^2 / \phi_1^4))^{1/2}}. \tag{10.11}$$

Furthermore,

$$\frac{n^{-1}X_1'X_1}{E(Y_0^{*2}/\phi_1^2)} \to_p 1, \ \frac{n^{-1}X_1'P_{X_2}X_1}{E(Y_0^{*2}/\phi_1^2)} \to_p 0, \ \frac{n^{-1}X_1'M_{X_2}\Delta^2 M_{X_2}X_1}{E(Y_0^{*2}U_1^2/\phi_1^4)} \to_p 1. \quad (10.12)$$

If $\rho \to \rho^* < 1$, we have

$$\frac{n^{-1}X_1'M_{X_2}X_1}{E(Y_0^{*2}/\phi_1^2) - (E(Y_0^*/\phi_1^2))^2/E(\phi_1^{-2})} \to_p 1. \quad (10.13)$$

AG-AR prove the following Lemma which is helpful in determining the order of the normalization sequences $a_n(\gamma_{n,h})$ and $d_n(\gamma_{n,h})$ in the case where $h = \infty$.

Lemma S1. *Suppose* $n(1-\rho) \to \infty$, $\rho \to 1$, *and Assumptions* INNOV *and* STAT *hold, then we have*

$$\begin{aligned} E(Y_0^{*2}U_1^2/\phi_1^4) - (1-\rho^2)^{-1}(EU_1^2)^2/\phi_1^4 &= O(1), \\ E(Y_0^{*2}/\phi_1^2) - (1-\rho^2)^{-1}EU_1^2 E\phi_1^{-2} &= O(1), \\ E(Y_0^*/\phi_1^2) &= O(1), \textit{ and} \\ E(Y_0^*U_1^2/\phi_1^4) &= O(1). \end{aligned}$$

A more detailed version of the following Lemma is proven in AG-AR as well.

Lemma S2 *Suppose Assumptions* INNOV *and* STAT *hold,* $\rho_n \in (-1,1]$, $\rho_n = 1 - h_{n,1}/n$ *where* $h_{n,1} \to h_1 \in (0,\infty)$. *Then, the following results* (a)-(c) *hold jointly,*
(a) $n^{-1}\sum_{i=1}^n \phi_{n,i}^{-j} \to_p \lim_{n\to\infty} E_{F_n}\phi_{n,i}^{-j} = h_{2,(j+3)}$ *for* $j = 1, 2, 4$,
(b) $n^{-1}\sum_{i=1}^n U_{n,i}^2/\phi_{n,i}^4 \to_p \lim_{n\to\infty} E_{F_n}(U_{n,i}^2/\phi_{n,i}^4) = h_{2,2}$, *and*
(c) $n^{-3/2}\sum_{i=1}^n Y_{n,i-1}^*/\phi_{n,i}^2 = O_p(1)$ *and* $n^{-3/2}\sum_{i=1}^n Y_{n,i-1}^*U_{n,i}^2/\phi_{n,i}^4 = O_p(1)$.

When $\rho_n = 1 - h_{n,1}/n$, where $h_{n,1} \to h_1 < \infty$, it is shown in AG-AR that

$$\begin{aligned} n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}\widehat{U}_i^2/\phi_i^4 &= n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}U_i^2/\phi_i^4 + o_p(1), \\ n^{-3/2}\sum_{i=1}^n Y_{i-1}^{*}\widehat{U}_i^2/\phi_i^4 &= n^{-3/2}\sum_{i=1}^n Y_{i-1}^{*}U_i^2/\phi_i^4 + o_p(1), \text{ and} \\ n^{-1}\sum_{i=1}^n \widehat{U}_i^2/\phi_i^4 &= n^{-1}\sum_{i=1}^n U_i^2/\phi_i^4 + o_p(1), \end{aligned} \quad (10.14)$$

where $\widehat{U}_i/\phi_i$ is the i-th residual from the LS regression of Y_i/ϕ_i on Y_{i-1}/ϕ_i and $1/\phi_i$.

10.2.4 Verification of Assumption E

In this section, we verify Assumption E for model 1. We make repeated use of the following well-known strong-mixing covariance inequality, see e.g. Doukhan (1994, Thm. 3, p. 9). Let X and Y be strong-mixing random variables with respect to sigma algebras $\mathcal{F}_i^j$ (for integers $i \leq j$) such that $X \in \mathcal{F}_{-\infty}^n$ and $Y \in \mathcal{F}_{n+k}^{\infty}$ with strong-mixing numbers $\{\alpha(k) : k \geq 1\}$. For $p, q > 0$ such that $1 - p^{-1} - q^{-1} > 0$, let $||X||_p = (E|X|^p)^{1/p}$ and $||Y||_q = (E|Y|^q)^{1/q}$. Then, the following inequality holds

$$Cov(X, Y) \leq 8||X||_p ||Y||_q \alpha(k)^{1-p^{-1}-q^{-1}}. \tag{10.15}$$

Below we apply the mixing inequality (10.15) with $p = q = \zeta > 3$, where ζ appears in Assumption INNOV. Assumption INNOV(iv) will imply that the expression $||X||_p ||Y||_q$ on the rhs of the inequality is $O(1)$.

For verification of Assumption E, as argued in the next paragraph, it is enough to show that for all $x \in R$, $U_{n,b_n}(x) - E_{\gamma_n} U_{n,b_n}(x) \rightarrow_p 0$ under $\{\gamma_n : n \geq 1\}$ for all sequences $\{\gamma_n = (1 - \rho_n, \gamma_{n,2}', \gamma_{n,3}')' \in \Gamma : n \geq 1\}$ that satisfy $n(1 - \rho_n) \rightarrow h_1$, $b(1 - \rho_n) \rightarrow g_1$, and $\gamma_{n,2} \rightarrow h_2 \in \Gamma_2$ for $(g, h) \in GH$, where $g = (g_1, h_2)$ and $h = (h_1, h_2)$.

To show $U_{n,b_n}(x) - E_{\gamma_n} U_{n,b_n}(x) \rightarrow_p 0$ under an arbitrary sequence $\{\gamma_n \in \Gamma : n \geq 1\}$ it is enough to show that for any subsequence $\{t_n\}$ there is a sub-subsequence $\{s_n\}$ such that $U_{s_n,b_{s_n}}(x) - E_{\gamma_{s_n}} U_{s_n,b_{s_n}}(x) \rightarrow_p 0$ under $\{\gamma_{s_n} \in \Gamma : n \geq 1\}$. Given any subsequence $\{t_n\}$ we can always construct a sub-subsequence $\{s_n\}$ such that $s_n(1 - \rho_{s_n}) \rightarrow h_1$, $b_{s_n}(1 - \rho_{s_n}) \rightarrow g_1$, and $\gamma_{s_n,2} \rightarrow h_2$ for $(g, h) \in GH$. Proceeding as in the proof of Lemma 6(iii) in AG1, we can define a sequence $\{\gamma_n^* : n \geq 1\}$ such that $n(1 - \rho_n^*) \rightarrow h_1$, $b(1 - \rho_n^*) \rightarrow g_1$, $\gamma_{n,2}^* \rightarrow h_2$, and $\gamma_{s_n}^* = \gamma_{s_n}$. It follows that $U_{n,b_n}(x) - E_{\gamma_n} U_{n,b_n}(x) \rightarrow_p 0$ holds under $\{\gamma_n^* : n \geq 1\}$ and therefore $U_{s_n,b_{s_n}}(x) - E_{\gamma_{s_n}} U_{s_n,b_{s_n}}(x) \rightarrow_p 0$ holds under $\{\gamma_{s_n} \in \Gamma : n \geq 1\}$.

For notational simplicity, in the rest of this section we let ρ denote ρ_n.

It is sufficient to show that for any given $x \in R$, $var(U_{n,b_n}(x)) \rightarrow 0$ under all sequences $\{\gamma_n \in \Gamma : n \geq 1\}$ that satisfy the conditions in the second paragraph of this subsection. Recall that $T_{n,b,k}(\rho)$ denotes the studentized t statistic based on the $k-$th subsample and the full-sample version is defined as $T_n^*(\theta_n) = n^{1/2}(\widehat{\rho} - \rho_n)/\widehat{\sigma}$, where $\widehat{\rho} = (X_1' M_{X_2} X_1)^{-1} X_1' M_{X_2} Y$, and $\widehat{\sigma}^2 = (n^{-1} X_1' M_{X_2} X_1)^{-2} (n^{-1} X_1' M_{X_2} \Delta^2 M_{X_2} X_1)$.[4] We write $T_{n,k}$ instead of $T_{n,b,k}(\rho)$ to simplify notation. Define

$$I_{b,k} = 1\{T_{n,k} \leq x\}. \tag{10.16}$$

Stationarity of $I_{b,k}$ in k implies that

$$var(U_{n,b_n}(x)) = q_n^{-1} var(I_{b,0}) + 2q_n^{-2} \sum_{k=1}^{q_n-1} (q_n - k) Cov(I_{b,0}, I_{b,k}). \tag{10.17}$$

[4] Here we deal with the upper one-sided case, so that $T_{n,b,k}(\rho) = T_{n,b,k}^*(\rho)$. The lower one-sided and symmetric two-sided cases can be dealt with using the same approach.

In this example, $q_n = n - b + 1$. Thus, it suffices to show $n^{-1} \sum_{k=0}^{n} |Cov(I_{b,0}, I_{b,k})| \to 0$. This is implied by

$$\sup_{k \geq k_n} |Cov(I_{b,0}, I_{b,k})| \to 0 \tag{10.18}$$

as $n \to \infty$ for some sequence $k_n \to \infty$ such that $k_n/n \to 0$.

By definition $T_{n,k} = b^{1/2}(\widehat{\rho}_{n,b,k} - \rho)/\widehat{\sigma}_{n,b,k}$. Below we show that for all $k \geq k_n$ we can write

$$T_{n,k} = \widetilde{T}_{n,k} + \eta_{n,k}, \tag{10.19}$$

for some random variables $\widetilde{T}_{n,k}$ and $\eta_{n,k}$ that are defined such that (i) $\widetilde{T}_{n,k}$ and $T_{n,0}$ are based on innovations, U_i, that are separated by at least b time periods and (ii) $\eta_{n,k} = o_p(1)$ uniformly in $k \geq k_n$ (by which we mean that $\forall \varepsilon > 0$, $\sup_{k \geq k_n} \Pr(|\eta_{n,k}| > \varepsilon) \to 0$). (Likewise, for an array $a_{n,k}$ of real numbers, we say that $a_{n,k}$ is $o(1)$ uniformly in $k \geq k_n$, if $\sup_{k \geq k_n} |a_{n,k}| \to 0$ as $n \to \infty$.) Note that the largest index of any U_i appearing in $T_{n,0}$ is $i = b$.

Using (10.19), we show below that

$$\begin{aligned} \sup_{k \geq k_n} |P(\widetilde{T}_{n,k} \leq x) - P(T_{n,k} \leq x)| \to 0 \text{ and} \\ \sup_{k \geq k_n} |P(T_{n,0} \leq x \ \&\ T_{n,k} \leq x) - P(T_{n,0} \leq x \ \&\ \widetilde{T}_{n,k} \leq x)| \to 0. \end{aligned} \tag{10.20}$$

Using these results, we obtain

$$\begin{aligned} Cov(I_{b,0}, I_{b,k}) &= EI_{b,0}I_{b,k} - EI_{b,0}EI_{b,k} \\ &= P(T_{n,0} \leq x \ \&\ T_{n,k} \leq x) - P(T_{n,0} \leq x)P(T_{n,k} \leq x) \\ &= P(T_{n,0} \leq x \ \&\ \widetilde{T}_{n,k} \leq x) - P(T_{n,0} \leq x)P(\widetilde{T}_{n,k} \leq x) + o(1) \\ &\leq \alpha(b) + o(1) \\ &= o(1), \end{aligned} \tag{10.21}$$

where the third equality holds by (10.20), the fourth equality holds by the definition of the α-mixing numbers $\{\alpha_n(m) : m = 1, 2, ...\}$ of $\{U_{n,i} : i = ..., 0, 1, ...\}$, where $\alpha(m) = \sup_{n \geq 1} \alpha_n(m)$, and the fact that $\widetilde{T}_{n,k}$ and $T_{n,0}$ are separated by at least b time periods, the last equality holds by the strong-mixing assumption in the definition of $\mathcal{F}(\gamma_2)$, and the $o(1)$ expression is uniform in $k \geq k_n$ by (10.20). Therefore (10.18) holds and the proof is complete except for the verifications of (10.19) and (10.20).

Equation (10.20) is established as follows. Equation (10.19) and $P(T_{n,k} \leq x) \to J_h(x)$ as $n \to \infty$ where J_h is continuous (see Proposition S1) imply that for all $\varepsilon > 0$ there exist $\delta > 0$ and $n_0 \in N$ such that for $n \geq n_0$ we have

$$\begin{aligned} &\sup_{k \geq k_n} P(|\widetilde{T}_{n,k} - T_{n,k}| > \delta) < \varepsilon/2, \\ &P(T_{n,k} \leq x + \delta) \leq P(T_{n,k} \leq x) + \varepsilon/2, \text{ and} \\ &P(T_{n,k} \leq x) \leq P(T_{n,k} \leq x - \delta) + \varepsilon/2. \end{aligned} \tag{10.22}$$

The latter two inequalities hold for all k because $T_{n,k}$ is identically distributed across k. These results lead to

$$
\begin{aligned}
&\sup_{k \geq k_n} |P(\widetilde{T}_{n,k} \leq x) - P(T_{n,k} \leq x)| \\
&= \sup_{k \geq k_n} \max\{P(\widetilde{T}_{n,k} \leq x) - P(T_{n,k} \leq x), -P(\widetilde{T}_{n,k} \leq x) + P(T_{n,k} \leq x)\} \\
&\leq \sup_{k \geq k_n} \max\{P(\widetilde{T}_{n,k} \leq x) - P(T_{n,k} \leq x + \delta), \\
&\quad -P(\widetilde{T}_{n,k} \leq x) + P(T_{n,k} \leq x - \delta)\} + \varepsilon/2 \\
&\leq \sup_{k \geq k_n} P(|\widetilde{T}_{n,k} - T_{n,k}| > \delta) + \varepsilon/2 \\
&\leq \varepsilon, \qquad (10.23)
\end{aligned}
$$

which proves the first result in (10.20). The second result in (10.20) can be proved in the same way. For example, the analogue of the third equation in (10.22) holds because

$$
\begin{aligned}
&P(T_{n,0} \leq x \ \&\ T_{n,k} \leq x) - P(T_{n,0} \leq x \ \&\ T_{n,k} \leq x - \delta) \\
&\leq P(x - \delta < T_{n,k} \leq x) = P(x - \delta < T_{n,0} \leq x) < \varepsilon/2 \qquad (10.24)
\end{aligned}
$$

for all k, for $\delta > 0$ small enough. This completes the proof of (10.20).

It remains to establish (10.19). We consider several cases. (i) $b(1-\rho) \to \infty$ with two subcases $\rho \to 1$ and $\rho \to \rho^* < 1$, (ii) $b(1-\rho) \to h_1 \in (0, \infty)$, (iii) $b(1-\rho) \to 0$ & $n(1-\rho) \to \infty$, and (iv) $n(1-\rho) \to h_1 \in [0, \infty)$.

Proof of (10.19) for case (i): $b(1-\rho) \to \infty$.

By Proposition S1, we know that for

$$
d_b = \frac{E_{F_n}(Y_0^{*2}/\phi_1^2) - (E_{F_n}(Y_0^*/\phi_1^2))^2 / E_{F_n}(\phi_1^{-2})}{(Z' E_{F_n}(X^1 X^{1\prime} U_1^2/\phi_1^2) Z)^{1/2}} = \frac{d_{b1}}{d_{b2}^{1/2}} \text{ and } a_b = b^{1/2} d_b, \qquad (10.25)
$$

we have $d_b \widehat{\sigma}_{n,b,k} \to_p 1$. Also, by (10.12) and (10.13) we have $d_{b1}^{-1} b^{-1} X_1' M_{X_2} X_1 \to_p 1$, where here (with abuse of notation) X_1 and X_2 denote b-vectors containing data from the k-th subsample. This implies that uniformly in k

$$
\begin{aligned}
T_{n,k} &= b^{-1/2} d_{b2}^{-1/2} \sum_{i=1}^{b} Y_{k+i-1}^* U_{k+i}/\phi_{k+i}^2 \qquad (10.26) \\
&- \left(b^{-1} d_{b2}^{-1/2} \sum_{j=1}^{b} Y_{k+j-1}^*/\phi_{k+j}^2 \right) \left(b^{-1} \sum_{j=1}^{b} \phi_{k+j}^{-2} \right)^{-1} b^{-1/2} \sum_{i=1}^{b} U_{k+i}/\phi_{k+i}^2 + o_p(1).
\end{aligned}
$$

Consider first the subcase where $\rho \to 1$. In that case, (10.11) implies further that uniformly in k

$$
T_{n,k} = b^{-1/2} d_{b2}^{-1/2} \sum_{i=1}^{b} Y_{k+i-1}^* U_{k+i}/\phi_{k+i}^2 + o_p(1). \qquad (10.27)
$$

Note that

$$\sum_{i=1}^{b} Y^*_{k+i-1} U_{k+i}/\phi^2_{k+i} = \sum_{i=1}^{b}\sum_{s=0}^{\infty} \rho^s U_{k+i-1-s} U_{k+i}/\phi^2_{k+i}. \tag{10.28}$$

Set

$$\widetilde{T}_{n,k} = b^{-1/2} d_{b2}^{-1/2} \sum_{i=1}^{b} \sum_{s=0}^{k+i-2b-2} \rho^s U_{k+i-1-s} U_{k+i}/\phi^2_{k+i} \tag{10.29}$$

and note that the smallest index of any U_i appearing in $\widetilde{T}_{n,k}$ is $i = 2b - 1$. We are just left with showing that

$$T_{n,k} - \widetilde{T}_{n,k} = b^{-1/2} d_{b2}^{-1/2} \sum_{i=1}^{b} \sum_{s=k+i-2b-1}^{\infty} \rho^s U_{k+i-1-s} U_{k+i}/\phi^2_{k+i} + o_p(1) = o_p(1). \tag{10.30}$$

To show (10.30), note that by Markov's inequality, we have

$$\begin{aligned}
&P(|b^{-1/2} d_{b2}^{-1/2} \sum_{i=1}^{b} \sum_{s=k+i-2b-1}^{\infty} \rho^s U_{k+i-1-s} U_{k+i}/\phi^2_{k+i}| > \varepsilon) \\
&\le \varepsilon^{-2} b^{-1} d_{b2}^{-1} \sum_{i,j=1}^{b} \sum_{\substack{s=k+i-2b-1 \\ t=k+j-2b-1}}^{\infty} \rho^{s+t} E U_{k+i-1-s} (U_{k+i}/\phi^2_{k+i}) U_{k+j-1-t} U_{k+j}/\phi^2_{k+j} \\
&= O(b^{-1}(1-\rho)) \sum_{i=1}^{b} \sum_{s,t=k+i-2b-1}^{\infty} \rho^{s+t} E U_{k+i-1-s} U_{k+i-1-t} U^2_{k+i}/\phi^4_{k+i},
\end{aligned} \tag{10.31}$$

where the equality holds by Lemma S1 and the fact that $E(U_{k+i-1-s}\ (U_{k+i}/\phi^2_{k+i}) \times U_{k+j-1-t} U_{k+j}/\phi^2_{k+j}) = 0$ for $i \neq j$. The contribution of all terms with $s = t$ is of order $o(1)$ because

$$\sum_{s=k+i-2b-1}^{\infty} \rho^{2s} = \sum_{s=0}^{\infty} \rho^{2(s+k+i-2b-1)} \le \rho^i \sum_{s=0}^{\infty} \rho^s = \rho^i (1-\rho)^{-1} \tag{10.32}$$

and $b^{-1}\sum_{i=1}^{b} \rho^i = o_p(1)$ since $b(1-\rho) \to \infty$. For the contributions with $s > t$, using (10.15) and Assumption INNOV(iv), the rhs of (10.31) equals

$$\begin{aligned}
&O(b^{-1}(1-\rho)) \sum_{i=1}^{b} \sum_{s>t=k+i-2b-1}^{\infty} \rho^{s+t} (s-t)^{-3-\varepsilon} \\
&= O(b^{-1}(1-\rho)) \sum_{i=1}^{b} \rho^i \sum_{s>t=0}^{\infty} \rho^{s+t} (s-t)^{-3-\varepsilon} \\
&= O(b^{-1}) \sum_{i=1}^{b} \rho^i \\
&= o(1),
\end{aligned} \tag{10.33}$$

where the first equality holds by the change of variables $s \to s + k + i - 2b - 1$ and similarly for t and the last equality uses $b(1-\rho) \to \infty$.

Next consider the subcase where $\rho \to \rho^* < 1$. In this case, define

$$\begin{aligned}
\widetilde{T}_{n,k} &= b^{-1/2} d_{b2}^{-1/2} \sum_{i=1}^{b} \sum_{s=0}^{k+i-2b-2} \rho^s U_{k+i-1-s} U_{k+i} / \phi_{k+i}^2 \\
&\quad - \left(b^{-1} d_{b2}^{-1/2} \sum_{j=1}^{b} \sum_{s=0}^{k+j-2b-2} \rho^s U_{k+j-1-s} / \phi_{k+j}^2 \right) \left(b^{-1} \sum_{j=1}^{b} \phi_{k+j}^{-2} \right)^{-1} \\
&\quad \times b^{-1/2} \sum_{i=1}^{b} U_{k+i} / \phi_{k+i}^2. \qquad (10.34)
\end{aligned}$$

Note that $(b^{-1} \sum_{j=1}^{b} \phi_{k+j}^{-2})^{-1}$, $b^{-1/2} \sum_{i=1}^{b} U_{k+i}/\phi_{k+i}^2$, and $d_{b2}^{-1/2}$ are all $O_p(1)$. (The first quantity is $O_p(1)$ by Lemma S2(a) and $h_{2,5} \geq \varepsilon_2 > 0$, the second quantity is $O_p(1)$ by a CLT, and the third quantity is $O_p(1)$ by Assumption INNOV(iv).) Therefore, it is enough to show that

$$\begin{aligned}
b^{-1/2} \sum_{i=1}^{b} \sum_{s=k+i-2b-1}^{\infty} \rho^s U_{k+i-1-s} U_{k+i} / \phi_{k+i}^2 &= o_p(1) \text{ and} \\
b^{-1} \sum_{i=1}^{b} \sum_{s=k+i-2b-1}^{\infty} \rho^s U_{k+i-1-s} / \phi_{k+i}^2 &= o_p(1). \qquad (10.35)
\end{aligned}$$

Using Markov's inequality we have

$$\begin{aligned}
& P(|b^{-1/2} \sum_{i=1}^{b} \sum_{s=k+i-2b-1}^{\infty} \rho^s U_{k+i-1-s} U_{k+i} / \phi_{k+i}^2| > \varepsilon) \\
&= O(b^{-1}) \sum_{i,j=1}^{b} \sum_{\substack{s=k+i-2b-1 \\ t=k+j-2b-1}}^{\infty} \rho^{s+t} E U_{k+i-1-s} U_{k+i} / \phi_{k+i}^2 U_{k+j-1-t} U_{k+j} / \phi_{k+j}^2 \\
&= O(b^{-1}) \sum_{i=1}^{b} \sum_{s,t=k+i-2b-1}^{\infty} \rho^{s+t} \\
&= O(b^{-1}) \sum_{i=1}^{b} \sum_{s,t=0}^{\infty} \rho^{s+t+2k+2i-4b-2} \\
&= O(b^{-1}) \sum_{i=1}^{b} \rho^{2i} \sum_{s,t=0}^{\infty} \rho^{s+t} \\
&= o(1), \qquad (10.36)
\end{aligned}$$

where in the second equality we use Assumption INNOV(iv). The second term in (10.35) can be handled analogously. This completes the proof for case (i).

For cases (ii)-(iv) we proceed as follows to establish (10.19). Define

$$\begin{aligned} c_k &= b^{-1} \sum_{j=1}^{b} \phi_{k+j}^{-2} \text{ and} \\ f_{k,i} &= Y_{k+i-1}^* - c_k^{-1} b^{-1} \sum_{j=1}^{b} Y_{k+j-1}^* / \phi_{k+j}^2. \end{aligned} \tag{10.37}$$

Note that

$$\begin{aligned} T_{n,k} &= b^{1/2}(\widehat{\rho}_{n,b,k} - \rho)/\widehat{\sigma}_{n,b,k} = S_{1,k}/S_{2,k}^{1/2}, \text{ where} \\ S_{1,k} &= b^{-1} \sum_{i=1}^{b} f_{k,i} U_{k+i}/\phi_{k+i}^2, \\ S_{2,k} &= b^{-2} \sum_{i=1}^{b} f_{k,i}^2 \widehat{U}_{k+i}^2/\phi_{k+i}^4, \end{aligned} \tag{10.38}$$

and $\widehat{U} = M_X U$.[5] We show below that $S_{1,k}$ and $S_{2,k}$ can be written as

$$\begin{aligned} & S_{1,k} = \widetilde{S}_{1,k} + \xi_{1,k} \text{ and } S_{2,k} = \widetilde{S}_{2,k} + \xi_{2,k}, \text{ where} \\ & \widetilde{S}_{1,k} \text{ and } \widetilde{S}_{2,k} \text{ are separated from } S_{1,0} \text{ and } S_{2,0} \text{ by } b \text{ time periods } \forall k \geq k_n, \\ & \xi_{1,k} = o_p(1), \text{ and } \xi_{2,k} = o_p(1). \end{aligned} \tag{10.39}$$

Note that

$$\begin{aligned} f_{k,i} &= \sum_{s=0}^{\infty} \rho^s U_{k+i-1-s} - c_k^{-1} b^{-1} \sum_{j=1}^{b} \sum_{s=0}^{\infty} \rho^s U_{k+j-1-s}/\phi_{k+j}^2 \\ &= \sum_{s=i-1}^{\infty} \rho^s U_{k+i-1-s} - c_k^{-1} b^{-1} \sum_{j=1}^{b} \sum_{s=j-1}^{\infty} \rho^s U_{k+j-1-s}/\phi_{k+j}^2 \\ &\quad + \sum_{s=0}^{i-2} \rho^s U_{k+i-1-s} - c_k^{-1} b^{-1} \sum_{j=1}^{b} \sum_{s=0}^{j-2} \rho^s U_{k+j-1-s}/\phi_{k+j}^2 \\ &= \rho^{i-1} \sum_{s=0}^{\infty} \rho^s U_{k-s} - c_k^{-1} b^{-1} \sum_{j=0}^{b-1} \rho^j \sum_{s=0}^{\infty} \rho^s U_{k-s}/\phi_{k+j+1}^2 \\ &\quad + \sum_{s=1}^{i-1} \rho^{-s+i-1} U_{k+s} - c_k^{-1} b^{-1} \sum_{s=0}^{b-2} \sum_{j=1}^{b-(s+1)} \rho^s U_{k+j}/\phi_{k+j+s+1}^2, \end{aligned} \tag{10.40}$$

[5] Strictly speaking, all sums over $i = 1, ..., b$ should be over $i = 1, ..., b-1$ because one observation from a block of length b is used as an initial observation given that lagged Y_i is a regressor. For notational simplicity, here and below, we sum to b rather than $b-1$.

where we used the transformation $s \mapsto -s+i-1$ for the first sum of the last row and changed the sequence of summation over j and s and applied the transformation $j \mapsto j+s$ in the second sum of the last row. Therefore, changing back the sequence of summation over j and s in the second sum, it follows that for $\rho < 1$

$$\begin{aligned} f_{k,i} &= a_{k,i}\sum_{j=0}^{\infty}\rho^j U_{k-j} + \sum_{j=1}^{b-1} c_{k,i,j}U_{k+j}, \text{ where} \\ a_{k,i} &= \rho^{i-1} - \frac{c_k^{-1}}{b}\sum_{l=0}^{b-1}\frac{\rho^l}{\phi_{k+l+1}^2} \text{ and} \\ c_{k,i,j} &= 1(j \le i-1)\rho^{i-j-1} - \frac{c_k^{-1}}{b}\sum_{l=0}^{b-(j+1)}\frac{\rho^l}{\phi_{k+j+l+1}^2}. \end{aligned} \tag{10.41}$$

Note that $a_{k,i}$ is random. When $\rho = 1$, (10.41) simplifies to

$$f_{k,i} = \sum_{j=1}^{b-1}\left(1(j \le i-1) - \frac{c_k^{-1}}{b}\sum_{l=0}^{b-(j+1)}\phi_{k+j+l+1}^{-2}\right)U_{k+j}. \tag{10.42}$$

By (10.38) and (10.45) below, (10.42) implies that $T_{n,k}$ is separated from $T_{n,0}$ by at least b time periods when $k > 2b$. Thus, if $\rho = 1$ for all n, (10.19) holds immediately. This leads us to only consider cases where $\rho < 1$ for all n. (Sequences in which $\rho = 1$ for some n and $\rho < 1$ for some n can be handled by analyzing subsequences.)

We now truncate the infinite sum in $f_{k,i}$ and for $k > 2b$ define

$$\begin{aligned} f_{k,i}^t &= a_{k,i}\sum_{j=0}^{k-2b-1}\rho^j U_{k-j} + \sum_{j=1}^{b-1} c_{k,i,j}U_{k+j} \\ &= Y_{k+i-1}^{*t} - c_k^{-1}b^{-1}\sum_{j=1}^{b} Y_{k+j-1}^{*t}/\phi_{k+j}^2, \text{ where} \\ Y_{l-1}^{*t} &= \sum_{s=0}^{l-2b-2}\rho^s U_{l-1-s}. \end{aligned} \tag{10.43}$$

Note that $f_{k,i}^t$ is obtained from $f_{k,i}$ by deleting all U_p with subindices $p < 2b+1$. Define

$$\begin{aligned} \widetilde{S}_{1,k} &= b^{-1}\sum_{i=1}^{b} f_{k,i}^t U_{k+i}/\phi_{k+i}^2, \\ \xi_{1,k} &= b^{-1}\sum_{i=1}^{b}(f_{k,i} - f_{k,i}^t)U_{k+i}/\phi_{k+i}^2 \\ &= b^{-1}\sum_{i=1}^{b} a_{k,i}(U_{k+i}/\phi_{k+i}^2)\sum_{j=k-2b}^{\infty}\rho^j U_{k-j}. \end{aligned} \tag{10.44}$$

For $k > 2b$, $\widetilde{S}_{1,k}$ depends only on innovations U_p for $p > 2b$ and $S_{1,0}$ and $S_{2,0}$ depend only on innovations U_p for $p \leq b$. Thus, for $k > 2b$, $\widetilde{S}_{1,k}$ is separated from $S_{1,0}$ and $S_{2,0}$ by at least b time periods.

Regarding $S_{2,k}$, note that by (10.14) and the definition in (10.37) we have

$$S_{2,k} = b^{-2} \sum_{i=1}^{b} f_{k,i}^2 U_{k+i}^2 / \phi_{k+i}^4 + o_p(1). \tag{10.45}$$

Set

$$\widetilde{S}_{2,k} = b^{-2} \sum_{i=1}^{b} (f_{k,i}^t)^2 U_{k+i}^2 / \phi_{k+i}^4. \tag{10.46}$$

For $k > 2b$, $\widetilde{S}_{2,k}$ depends only on innovations U_p for $p > 2b$. By definition,

$$\xi_{2,k} = b^{-2} \sum_{i=1}^{b} (f_{k,i}^2 - (f_{k,i}^t)^2) U_{k+i}^2 / \phi_{k+i}^4 + o_p(1). \tag{10.47}$$

We now show that $\xi_{1,k} = o_p(1)$ and $\xi_{2,k} = o_p(1)$ uniformly for $k \geq k_n$ for some sequence $k_n \to \infty$ such that $k_n/n \to 0$.

Case (ii): $b(1-\rho) \to h_1 \in (0, \infty)$.

We first show that $\xi_{1,k} = o_p(1)$. Clearly, it is enough to show that

$$\begin{aligned} & b^{-1} \sum_{i=1}^{b} \rho^{i-1} U_{k+i} / \phi_{k+i}^2 \sum_{j=k-2b}^{\infty} \rho^j U_{k-j} = o_p(1) \text{ and} \\ & ((c_k^{-1}/b) \sum_{l=0}^{b-1} \rho^l \phi_{k+l+1}^{-2})(b^{-1} \sum_{i=1}^{b} U_{k+i}/\phi_{k+i}^2)(\sum_{j=k-2b}^{\infty} \rho^j U_{k-j}) = o_p(1). \end{aligned} \tag{10.48}$$

Note that by Lemma S2(a), c_k^{-1} and $b^{-1} \sum_{l=0}^{b-1} \rho^l \phi_{k+l+1}^{-2}$ are both $O_p(1)$. Applying Markov's inequality, it is therefore enough to show that the following quantity is $o_p(1)$:

$$\begin{aligned} & b^{-2} \sum_{i,j=1}^{b} \sum_{l,m=k-2b}^{\infty} \rho^{l+m} E(U_{k+i}/\phi_{k+i}^2)(U_{k+j}/\phi_{k+j}^2) U_{k-l} U_{k-m} \\ = & O(1) b^{-2} \sum_{i=1}^{b} \sum_{l,m=k-2b}^{\infty} \rho^{l+m} E U_{k-l} U_{k-m} U_{k+i}^2 / \phi_{k+i}^4, \end{aligned} \tag{10.49}$$

where the equality uses $E(U_{k+i}/\phi_{k+i}^2)(U_{k+j}/\phi_{k+j}^2)U_{k-l}U_{k-m} = 0$ for $k > 2b$ unless $i = j$, by the martingale difference property of U_i.

Consider first the contribution of the summands in (10.49) when $l = m$:

$$b^{-2}\sum_{i=1}^{b}\sum_{l=k-2b}^{\infty}\rho^{2l}EU_{k-l}^2U_{k+i}^2/\phi_{k+i}^4 = O(b^{-2})\sum_{i=1}^{b}\sum_{l=k-2b}^{\infty}\rho^{2l} = O(\rho^{k-2b}(b(1-\rho))^{-1}), \tag{10.50}$$

where in the first equality we use Assumption INNOV(iv). Define $h_{n,1}^*$ by $\rho = \exp(-h_{n,1}^*/n)$. Because $b(1-\rho) \to h_1 \in (0,\infty)$, we have $h_{n,1}^* \to \infty$. In consequence, there exists a sequence $\{k_n : n \geq 1\}$ such that $k_n/b \to \infty$, $k_n/n \to 0$ and $h_{n,1}^* k_n/n \to \infty$. For this sequence, $h_{n,1}^*(k_n - 2b)/n \to \infty$, $\rho^{k_n-2b} = \exp(-h_{n,1}^*(k_n - 2b)/n) \to 0$, and $\sup_{k\geq k_n}\rho^{2(k-2b)} \to 0$. This shows that the expression in (10.50) is $o(1)$.

Therefore, we only need to consider the contributions in (10.49) with $l > m$. We have

$$\begin{aligned}
& b^{-2}\sum_{i=1}^{b}\sum_{l>m=k-2b}^{\infty}\rho^{l+m}EU_{k-l}U_{k-m}U_{k+i}^2/\phi_{k+i}^4 \\
&= O(1)b^{-2}\sum_{i=1}^{b}\sum_{l>m=k-2b}^{\infty}\rho^{l+m}(l-m)^{-3-\varepsilon} \\
&= O(1)\rho^{k-2b}b^{-2}\sum_{i=1}^{b}\sum_{l>m=0}^{\infty}\rho^{m}(l-m)^{-3-\varepsilon} \\
&= o(1),
\end{aligned} \tag{10.51}$$

where in the first equality we use (10.15) and Assumption INNOV(iv).

Next we show $\xi_{2,k} = o_p(1)$. Note that up to a $o_p(1)$ term $\xi_{2,k} = \xi_{21,k} - 2\xi_{22,k} + \xi_{23,k}$, where

$$\begin{aligned}
\xi_{21,k} &= b^{-2}\sum_{i=1}^{b}(Y_{k+i-1}^{*2} - Y_{k+i-1}^{*t2})U_{k+i}^2/\phi_{k+i}^4, \\
\xi_{22,k} &= c_k^{-1}b^{-3/2}\sum_{i=1}^{b}\left(Y_{k+i-1}^*\left(b^{-3/2}\sum_{j=1}^{b}Y_{k+j-1}^*/\phi_{k+j}^2\right) - \right. \\
&\qquad \left. Y_{k+i-1}^{*t}\left(b^{-3/2}\sum_{j=1}^{b}Y_{k+j-1}^{*t}/\phi_{k+j}^2\right)\right)U_{k+i}^2/\phi_{k+i}^4, \text{ and} \\
\xi_{23,k} &= \left(\left(b^{-3/2}\sum_{j=1}^{b}Y_{k+j-1}^*/\phi_{k+j}^2\right)^2 - \left(b^{-3/2}\sum_{j=1}^{b}Y_{k+j-1}^{*t}/\phi_{k+j}^2\right)^2\right) \\
&\qquad \times c_k^{-2}b^{-1}\sum_{i=1}^{b}U_{k+i}^2/\phi_{k+i}^4.
\end{aligned} \tag{10.52}$$

To show $\xi_{21,k} = o_p(1)$, note that

$$\xi_{21,k} = b^{-2} \sum_{i=1}^{b} \sum_{\substack{s,t=0, \\ s \text{ or } t \geq k+i-2b-1}}^{\infty} \rho^{s+t} U_{k+i-1-s} U_{k+i-1-t} U_{k+i}^2 / \phi_{k+i}^4 \tag{10.53}$$

(where the second sum is over all $s, t = 0, \ldots$ for which $s \geq k+i-2b-1$ or $t \geq k+i-2b-1$). By Markov's inequality, we have

$$\begin{aligned} & P(|\xi_{21,k}| > \varepsilon) \\ \leq\ & \varepsilon^{-2} b^{-4} \sum_{i,j=1}^{b} \sum_{\substack{s,t=0, \\ s \text{ or } t \geq k+i-2b-1}}^{\infty} \sum_{\substack{u,v=0, \\ u \text{ or } v \geq k+i-2b-1}}^{\infty} \rho^{s+t+u+v} \\ & \times E(U_{k+i}^2/\phi_{k+i}^4)(U_{k+j}^2/\phi_{k+j}^4) U_{k+i-1-s} U_{k+i-1-t} U_{k+j-1-u} U_{k+j-1-v}. \end{aligned} \tag{10.54}$$

Using (10.15), Assumption INNOV(iv), and $\rho^{k-2b} \to 0$, one can show that the contribution of all summands for which at least two of the indices $k+i-1-s$, $k+i-1-t$, $k+j-1-u$, $k+j-1-v$ coincide is of order $o(1)$. In what follows, we can therefore assume that these indices are all different. We can then assume $i \geq j$, $s > t$, and $u > v$. One has to separately investigate several cases regarding the order of the six indices $k+i-1-s < k+i-1-t < k+i$ and $k+j-1-u < k+j-1-v < k+j$. We will only deal with the case where in the ordering of the indices ($k+i-1-s, k+i-1-t$, $k+j-1-u, k+j-1-v$, $k+j, k+i$) the subindex $k+i-1-s$ is followed immediately by $k+i-1-t$, the subindex $k+j-1-u$ is directly followed by $k+j-1-v$, $k+i-1-s \neq k+j$, and $k+j-1-u \neq k+i$. The other cases are dealt with analogously. Equation (10.15) and Assumption INNOV(iv) yield

$$\begin{aligned} & EU_{k+i-1-s} U_{k+i-1-t} U_{k+j-1-u} U_{k+j-1-v} (U_{k+j}^2/\phi_{k+j}^4) U_{k+i}^2/\phi_{k+i}^4 \\ \leq\ & O(\max\{s-t, u-v\})^{-3-\varepsilon}. \end{aligned} \tag{10.55}$$

Therefore, the summands in (10.54) equal

$$\begin{aligned} & O(b^{-4}) \sum_{i \geq j=1}^{b} \sum_{\substack{s>t=0, \\ s \geq k+i-2b-1}}^{\infty} \sum_{\substack{u>v=0, \\ u \geq k+i-2b-1}}^{\infty} \rho^{s+t+u+v} (\max\{s-t, u-v\})^{-3-\varepsilon} \\ = & O(b^{-4} \rho^{k-2b}) \sum_{i \geq j=1}^{b} \sum_{\substack{s>t=0, \\ s \geq k+i-2b-1}}^{\infty} \rho^t (s-t)^{\frac{-3-\varepsilon}{2}} \sum_{\substack{u>v=0, \\ u \geq k+i-2b-1}}^{\infty} \rho^v (u-v)^{\frac{-3-\varepsilon}{2}} \\ = & o(b^{-2}) \left(\sum_{t=0}^{\infty} \rho^t \sum_{s=t+1}^{\infty} (s-t)^{\frac{-3-\varepsilon}{2}} \right)^2. \end{aligned} \tag{10.56}$$

By a change of variable $s \to s+t+1$, the rhs of (10.56) equals

$$o(b^{-2})\left(\sum_{t=0}^{\infty}\rho^t\sum_{s=1}^{\infty}s^{\frac{-3-\varepsilon}{2}}\right)^2 = o(b^{-2})\left(\sum_{t=0}^{\infty}\rho^{2t}\right)^2 = o(b^{-2}(1-\rho)^{-2}) = o(1). \quad (10.57)$$

Next we deal with $\xi_{22,k}$. Note that by Lemma S2(a) and (c) we have $c_k^{-1} = O_p(1)$ and $b^{-3/2}\sum_{j=1}^{b} Y^*_{k+j-1}/\phi^2_{k+j} = O_p(1)$. We add and subtract $Y^{*t}_{k+i-1}b^{-3/2}\sum_{j=1}^{b} Y^*_{k+j-1}$ $/\phi^2_{k+j}$ which implies that it is enough to show that

$$\begin{aligned} b^{-3/2}\sum_{i=1}^{b}(Y^*_{k+i-1} - Y^{*t}_{k+i-1})U^2_{k+i}/\phi^4_{k+i} &= o_p(1), \\ b^{-3/2}\sum_{j=1}^{b}(Y^*_{k+j-1} - Y^{*t}_{k+j-1})/\phi^2_{k+j} &= o_p(1), \text{ and} \\ b^{-3/2}\sum_{i=1}^{b}Y^{*t}_{k+i-1}U^2_{k+i}/\phi^4_{k+i} &= O_p(1). \end{aligned} \quad (10.58)$$

The third statement holds by the first one and by Lemma S2(c). To prove the first two statements, note that

$$Y^*_{k+i-1} - Y^{*t}_{k+i-1} = \sum_{s=k+i-2b-1}^{\infty}\rho^s U_{k+i-1-s}. \quad (10.59)$$

To show (10.58), by Markov's inequality it is sufficient to show that

$$\begin{aligned} b^{-3}\sum_{i,j=1}^{b}\sum_{\substack{s=k+i-2b-1,\\ t=k+i-2b-1}}^{\infty}\rho^{s+t}EU_{k+i-1-s}U_{k+j-1-t}(U^2_{k+i}/\phi^4_{k+i})U^2_{k+j}/\phi^4_{k+j} &= o(1) \text{ and} \\ b^{-3}\sum_{i,j=1}^{b}\sum_{\substack{s=k+i-2b-1,\\ t=k+i-2b-1}}^{\infty}\rho^{s+t}EU_{k+i-1-s}U_{k+j-1-t}\phi^{-2}_{k+i}\phi^{-2}_{k+j} &= o(1). \end{aligned} \quad (10.60)$$

These can be shown using the method employed above. Finally, $\xi_{23,k} = o_p(1)$ follows by similar steps to the ones above and Lemma S2(b). This completes the proof of case (ii).

Case (iii) $b(1-\rho) \to 0$ **&** $n(1-\rho) \to \infty$.

Define $h^*_{n,1}$ and $h_{n,1}$ by $\rho = \exp(-h^*_{n,1}/n)$ and $\rho = 1 - h_{n,1}/n$. Let $t_n = bh^*_{n,1}/n$. For notational simplicity, we write $h^*_{n,1}$ and $h_{n,1}$ as h^*_n and h_n, respectively, in the remainder of the verification of Assumption E. Then, we have

$$\begin{aligned} &\rho^b = \exp(-bh^*_n/n) = \exp(-t_n),\ 1+\rho = 2 - h_n/n, \text{ and} \\ &b(1-\rho) = bh_n/n = t_n(h_n/h^*_n). \end{aligned} \quad (10.61)$$

We have: $b(1-\rho) \to 0 \Rightarrow \rho \to 1 \Rightarrow h_n^*/n \to 0 \Rightarrow h_n^*/h_n \to 1$, where the last implication follows from a mean-value expansion of $\exp(-h_n^*/n)$ about 0. In addition, $b(1-\rho) \to 0 \Rightarrow bh_n/n \to 0$. Combining these results gives $t_n = (bh_n/n)(h_n^*/h_n) \to 0$. Also, $n(1-\rho) \to \infty$ implies that $h_n \to \infty$ and $h_n^* \to \infty$.

Because $bh_n/n = b(1-\rho) \to 0$ it follows that $h_n = o(n/b)$. This and $h_n^*/h_n \to 1$ yield $h_n^* = o(n/b)$. By an expansion of $\exp(-h_n^*/n)$ about 0, we obtain

$$\begin{aligned} 0 &= \rho - \rho = \exp(-h_n^*/n) - (1 - h_n/n) \\ &= -h_n^*/n + 2^{-1}(h_n^*/n)^2 - 6^{-1}\exp(-h_n^{**}/n)(h_n^*/n)^3 + h_n/n, \end{aligned} \tag{10.62}$$

where $h_n^{**}/n = o(1/b)$ because $h_n^* = o(n/b)$. Hence,

$$1 - h_n/h_n^* = 2^{-1}h_n^*/n - 6^{-1}(h_n^*/n)^2 \exp(-h_n^{**}/n). \tag{10.63}$$

We first verify (10.39) for $\xi_{1,k} = b^{-1}\sum_{i=1}^{b} a_{k,i}(U_{k+i}/\phi_{k+i}^2)\sum_{j=k-2b}^{\infty}\rho^j U_{k-j}$ defined in (10.44). Note that by Markov's inequality we have

$$P(|\sum_{s=k-2b}^{\infty}\rho^s U_{k-s}| > M(1-\rho)^{-1/2}) \leq M^{-2}(1-\rho)\sum_{s=k-2b}^{\infty}\rho^{2s}EU_{k-s}^2 = O(M^{-2}) \tag{10.64}$$

by Assumption INNOV(iv) and because U_i is a martingale difference sequence. Therefore

$$(1-\rho)^{1/2}\sum_{s=k-2b}^{\infty}\rho^s U_{k-s} = O_p(1). \tag{10.65}$$

To show $\xi_{1,k} = o_p(1)$, it is thus sufficient to show that

$$\zeta_1 = (1-\rho)^{-1/2}b^{-1}\sum_{i=1}^{b} a_{k,i}(U_{k+i}/\phi_{k+i}^2) = o_p(1). \tag{10.66}$$

By adding and subtracting 1, we can write

$$a_{k,i} = (\rho^{i-1} - 1) - c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l - 1)\phi_{k+l+1}^{-2}. \tag{10.67}$$

Therefore

$$\zeta_1 = (1-\rho)^{-1/2}b^{-1}\sum_{i=1}^{b}\left((\rho^{i-1}-1) - c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l - 1)\phi_{k+l+1}^{-2}\right)(U_{k+i}/\phi_{k+i}^2) \tag{10.68}$$

and it is enough to show that

$$\zeta_{11} = (1-\rho)^{-1/2} b^{-1} \sum_{i=1}^{b} (\rho^{i-1}-1)(U_{k+i}/\phi_{k+i}^2) = o_p(1) \text{ and}$$
$$\zeta_{12} = (1-\rho)^{-1/2} b^{-1/2} \left(c_k^{-1} b^{-1} \sum_{l=0}^{b-1} (\rho^l - 1)\phi_{k+l+1}^{-2} \right) b^{-1/2} \sum_{i=1}^{b} (U_{k+i}/\phi_{k+i}^2) = o_p(1). \tag{10.69}$$

To show $\zeta_{11} = o_p(1)$, by Markov's inequality, it is enough to show that

$$(1-\rho)^{-1} b^{-2} \sum_{i=1}^{b} (\rho^{i-1}-1)^2 = o(1), \tag{10.70}$$

where we use the fact that U_{k+i}/ϕ_{k+i}^2 is a martingale difference sequence and $E(U_{k+i}^2/\phi_{k+i}^4)$ is uniformly bounded by Assumption INNOV(iv). Writing the sum in (10.70) in closed form, it follows that it is enough to show that

$$\frac{1-\rho^{2b} - 2(1-\rho^b)(1+\rho) + b(1-\rho)(1+\rho)}{b^2(1-\rho)^2} = o(1). \tag{10.71}$$

Using (10.61) and (10.63) the lhs of (10.71) equals

$$\frac{1-\exp(-2t_n) - 2(1-\exp(-t_n))(1+\rho) + t_n(h_n/h_n^*)(1+\rho)}{(t_n(h_n/h_n^*))^2}. \tag{10.72}$$

We first show that replacing $(1+\rho)$ by 2 and (h_n/h_n^*) by 1 in (10.72) is negligible in the sense that

$$t_n^{-2}[-2(1-\exp(-t_n))(1+\rho-2) + t_n((h_n/h_n^*)(1+\rho)-2)] = o(1). \tag{10.73}$$

To show (10.73), note that by (10.63) $h_n/h_n^* = 1 - 2^{-1}h_n^*/n + 6^{-1}(h_n^*/n)^2 \exp(-h_n^{**}/n)$, where $h_n^{**}/n \to 0$. By a Taylor expansion for $\rho = \exp(-h_n^*/n)$ we have $\rho - 1 = -h_n^*/n + 2^{-1}(h_n^*/n)^2 \exp(-h_n^{++}/n)$ for some h_n^{++} such that $h_n^{++}/n \to 0$. By a Taylor expansion for $\exp(-t_n)$ we have $1 - \exp(-t_n) = t_n - 2^{-1}t_n^2 \exp(t_n^*)$ for some t_n^* such that $t_n^* \to 0$. Multiplying out, shows that the lhs in (10.73) is of order $t_n^{-2}(O(t_n(h_n^*/n)^2) + O(t_n^2 h_n^*/n))$ which is $o(1)$.

By applying l'Hopital's rule twice, the limit of the expression in (10.72) with $(1+\rho)$ replaced by 2 and (h_n/h_n^*) replaced by 1 equals 0 which completes the proof of $\zeta_{11} = o_p(1)$.

To show $\zeta_{12} = o_p(1)$, a CLT for martingale difference sequences shows that $b^{-1/2}\sum_{i=1}^{b} (U_{k+i}/\phi_{k+i}^2) = O_p(1)$. Furthermore, by Assumption INNOV(v) and (vii) we have $c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)\phi_{k+l+1}^{-2} = O_p(1)b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)$ and it is therefore enough to show that

$$\begin{aligned}(1-\rho)^{-1/2}b^{-1/2}b^{-1}\sum_{l=0}^{b-1}(\rho^l-1) &= o(1) \text{ or}\\ b^{-3/2}(1-\rho)^{-3/2}(1-\rho^b-b(1-\rho)) &= o(1). \qquad (10.74)\end{aligned}$$

Using analogous steps as in the proof for $\zeta_{11}=o_p(1)$ above then shows $\zeta_{12}=o_p(1)$. This completes the verification of (10.39) for $\xi_{1,k}$.

We are left with showing that the component $b^{-2}\sum_{i=1}^{b}(f_{k,i}^2-(f_{k,i}^t)^2)U_{k+i}^2/\phi_{k+i}^4$ of $\xi_{2,k}$ in (10.47) is of order $o_p(1)$. Using the definitions of $f_{k,i}$ and $f_{k,i}^t$ in (10.41) and (10.43), it follows that

$$\begin{aligned}&f_{k,i}^2-(f_{k,i}^t)^2\\ &= a_{k,i}^2\left(\sum_{j=0}^{\infty}\rho^j U_{k-j}\right)^2 - a_{k,i}^2\left(\sum_{j=0}^{k-2b-1}\rho^j U_{k-j}\right)^2\\ &\quad +2\sum_{j=1}^{b-1}c_{k,i,j}U_{k+j}\left(a_{k,i}\sum_{s=k-2b}^{\infty}\rho^s U_{k-s}\right)\\ &= a_{k,i}^2\sum_{\substack{j,\ell=0,\\ j \text{ or } \ell=\geq k-2b}}^{\infty}\rho^{j+l}U_{k-j}U_{k-l} + 2\sum_{s=k-2b}^{\infty}\rho^s U_{k-s}\sum_{j=1}^{b-1}a_{k,i}c_{k,i,j}U_{k+j}\\ &= f_{1,k,i}+f_{2,k,i}. \qquad (10.75)\end{aligned}$$

We first show that the contributions of $f_{1,k,i}$ to $\xi_{2,k}$ are of order $o_p(1)$. Note that

$$\begin{aligned}&b^{-2}\sum_{i=1}^{b}f_{1,k,i}U_{k+i}^2/\phi_{k+i}^4\\ &= \sum_{\substack{j,\ell=0,\\ j \text{ or } \ell=\geq k-2b}}^{\infty}\rho^{j+l}U_{k-j}U_{k-l}b^{-2}\sum_{i=1}^{b}a_{k,i}^2U_{k+i}^2/\phi_{k+i}^4\\ &= O_p((1-\rho)^{-1})b^{-2}\sum_{i=1}^{b}a_{k,i}^2U_{k+i}^2/\phi_{k+i}^4. \qquad (10.76)\end{aligned}$$

Using (10.67), it is therefore enough to show that

$$(1-\rho)^{-1}b^{-2}\sum_{i=1}^{b}(\rho^{i-1}-1)^2U_{k+i}^2/\phi_{k+i}^4 = o_p(1),$$

$$(1-\rho)^{-1}b^{-2}\left(c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)\phi_{k+l+1}^{-2}\right)\sum_{i=1}^{b}(\rho^{i-1}-1)U_{k+i}^2/\phi_{k+i}^4=o_p(1),$$
$$(1-\rho)^{-1}b^{-2}\left(c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)\phi_{k+l+1}^{-2}\right)^2\sum_{i=1}^{b}U_{k+i}^2/\phi_{k+i}^4=o_p(1). \tag{10.77}$$

To deal with the first term, it is enough to show that the LLN $b^{-1}\sum_{i=1}^{b}Z_{bi}=O_p(1)$ applies with $Z_{bi}=(b(1-\rho))^{-2}\,(\rho^{i-1}-1)^2U_{k+i}^2/\phi_{k+i}^4$. The LLN holds by White (1984, Theorem 3.47 with $r=\delta=3/2$) because Z_{bi} is α-mixing of size 3, has finite mean by Assumption INNOV(iv) and because $(\rho^b-1)(b(1-\rho))^{-1}=O(1)$, and because $\sum_{i=1}^{\infty}(i^{-3}E|Z_{bi}-EZ_{bi}|^3)^{2/3}<\infty$. The latter holds because by $(\rho^b-1)(b(1-\rho))^{-1}=O(1)$ and Assumption INNOV(iv), $E|Z_{bi}-EZ_{bi}|^3$ is uniformly bounded.

The proofs for the second and third terms in (10.77) are analogous. Just note that $c_k^{-1}=O_p(1)$ by Lemma S2(a) and that

$$b^{-1}\sum_{l=0}^{b-1}Z_{bl}^*=O_p(1) \tag{10.78}$$

applies also with $Z_{bl}^*=(b(1-\rho))^{-1}\,(\rho^l-1)\phi_{k+l+1}^{-2}$, $Z_{bl}^*=(b(1-\rho))^{-1}\,(\rho^l-1)U_{k+i}^2/\phi_{k+i}^4$, and $Z_{bl}^*=U_{k+i}^2/\phi_{k+i}^4$ by White (1984, Theorem 3.47 with $r=\delta=3/2$).

We next show that the contributions of $f_{2,k,i}$ to $\xi_{2,k}$ are of order $o_p(1)$. By (10.65) and (10.75) it is sufficient to show that

$$(1-\rho)^{-1/2}b^{-2}\sum_{j=1}^{b-1}\sum_{i=1}^{b}a_{k,i}c_{k,i,j}(U_{k+i}^2/\phi_{k+i}^4)U_{k+j}=o_p(1). \tag{10.79}$$

By replacing $a_{k,i}$ and $c_{k,i,j}$ by their definitions we have

$$\sum_{j=1}^{b-1}\sum_{i=1}^{b}a_{k,i}c_{k,i,j}(U_{k+i}^2/\phi_{k+i}^4)U_{k+j}=\sum_{j=1}^{b-1}\sum_{i=1}^{b}\left((\rho^{i-1}-1)-c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)\phi_{k+l+1}^{-2}\right)$$
$$\times\left(1(j\le i-1)\rho^{i-j-1}-\frac{c_k^{-1}}{b}\sum_{l=0}^{b-(j+1)}\frac{\rho^l}{\phi_{k+j+l+1}^2}\right)(U_{k+i}^2/\phi_{k+i}^4)U_{k+j}. \tag{10.80}$$

Multiplying out in (10.80), it is clear that in order to show (10.79), it is sufficient to show that the following expressions multiplied by $(1-\rho)^{-1/2}b^{-2}$ are all $o_p(1)$:

$$\sum_{j=1}^{b-1}\sum_{i=1}^{b}(\rho^{i-1}-1)1(j\le i-1)\rho^{i-j-1}(U_{k+i}^2/\phi_{k+i}^4)U_{k+j},$$

$$\sum_{j=1}^{b-1}\sum_{i=1}^{b} c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)\phi_{k+l+1}^{-2}1(j\leq i-1)\rho^{i-j-1}(U_{k+i}^2/\phi_{k+i}^4)U_{k+j},$$

$$\left(\sum_{i=1}^{b}(\rho^{i-1}-1)(U_{k+i}^2/\phi_{k+i}^4)\right)\sum_{j=1}^{b-1}\left(\frac{c_k^{-1}}{b}\sum_{l=0}^{b-(j+1)}\frac{\rho^l}{\phi_{k+j+l+1}^2}\right)U_{k+j},$$

$$\left(c_k^{-1}b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)\phi_{k+l+1}^{-2}\right)\left(\sum_{j=1}^{b-1}\frac{c_k^{-1}}{b}\sum_{l=0}^{b-(j+1)}\frac{\rho^l}{\phi_{k+j+l+1}^2}U_{k+j}\right)\sum_{i=1}^{b}(U_{k+i}^2/\phi_{k+i}^4). \tag{10.81}$$

From the LLN in (10.78) and from $c_k^{-1}=O_p(1)$ it follows that in order to show that the expressions in (10.81) multiplied by $(1-\rho)^{-1/2}b^{-2}$ are $o_p(1)$ it is sufficient to show that

$$(1-\rho)^{1/2}b^{-1}\sum_{j=1}^{b-1}\sum_{i=j+1}^{b}\frac{\rho^{i-1}-1}{b(1-\rho)}\rho^{i-j-1}(U_{k+i}^2/\phi_{k+i}^4)U_{k+j}=o_p(1),$$

$$(1-\rho)^{1/2}b^{-1}\sum_{j=1}^{b-1}\sum_{i=j+1}^{b}\rho^{i-j-1}(U_{k+i}^2/\phi_{k+i}^4)U_{k+j}=o_p(1),$$

$$(1-\rho)^{1/2}b^{-1}\sum_{j=1}^{b-1}\sum_{l=0}^{b-(j+1)}\rho^l\phi_{k+j+l+1}^{-2}U_{k+j}=o_p(1). \tag{10.82}$$

To see this, note that the first row in (10.81) is clearly implied by the first row in (10.82). The second row in (10.81) is implied by the second row in (10.82) because in (10.81) we apply the LLN in (10.78) to $b^{-1}\sum_{l=0}^{b-1}(\rho^l-1)\phi_{k+l+1}^{-2}$ which is thus of order $O_p(b(1-\rho))$. The same LLN argument applied to $b^{-1}\sum_{i=1}^{b}(\rho^{i-1}-1)(U_{k+i}^2/\phi_{k+i}^4)$ shows that the third row in (10.81) is implied by the third row in (10.82). The fourth row in (10.81) is implied by the previous arguments and $b^{-1}\sum_{i=1}^{b}(U_{k+i}^2/\phi_{k+i}^4)=O_p(1)$.

For the third term in (10.82), by Markov's inequality, it is enough to show that

$$(1-\rho)b^{-2}\sum_{j=1}^{b-1}\sum_{l=0}^{b-(j+1)}\sum_{i=1}^{b-1}\sum_{m=0}^{b-(i+1)}\rho^{l+m}E\phi_{k+j+l+1}^{-2}U_{k+j}\phi_{k+i+m+1}^{-2}U_{k+i}=o(1). \tag{10.83}$$

We can assume that $i\neq j$ because the contributions of all summands with $i=j$ can be bounded by $(1-\rho)b^{-2}\sum_{i=1}^{b-1}\sum_{l,m=0}^{b-(i+1)}E\phi_{k+j+l+1}^{-2}\phi_{k+i+m+1}^{-2}U_{k+i}^2$ which is $o(1)$ because $E\phi_{k+j+l+1}^{-2}\phi_{k+i+m+1}^{-2}U_{k+i}^2$ is uniformly bounded by Assumption INNOV(iv) and $(1-\rho)b^{-2}b^3=o(1)$. Using the same argument we can assume that all subindices $k+j+l+1$, $k+i+m+1$, $k+j$, and $k+i$ are different and also that $i>j$. We have to distinguish two subcases, namely $k+j+l+1>k+i$ and $k+j+l+1<k+i$. The contributions of all summands in the lhs of (10.83) satisfying $k+j+l+1>k+i$

can be bounded by

$$O(1)(1-\rho)b^{-2}\sum_{i>j=1}^{b-1}\sum_{l=0}^{b-(j+1)}\sum_{m=0,m\neq l}^{b-(i+1)}(i-j)^{-3-\varepsilon}=O(1-\rho)\sum_{i>j=1}^{b-1}(i-j)^{-3-\varepsilon}=o(1), \tag{10.84}$$

where the first expression uses the strong mixing inequality (10.15) and Assumption INNOV(iv) and the last equality uses $b(1-\rho)\to 0$. The contributions of all summands in the lhs of (10.83) satisfying $k+j+l+1\leq k+i$ can be bounded by

$$O(1)(1-\rho)b^{-2}\sum_{i>j=1}^{b-1}\sum_{l=0}^{b-(j+1)}\sum_{m=0,m\neq l}^{b-(i+1)}(m+1)^{-3-\varepsilon}=O(1-\rho)b\sum_{m=0}^{b}(m+1)^{-3-\varepsilon}=o(1). \tag{10.85}$$

The first and second term in (10.82) are handled in exactly the same way. For the first term, recall that $(b(1-\rho))^{-1}(\rho^{i-1}-1)$ is $O(1)$ uniformly in i.

Case (iv) $n(1-\rho)\to h_1\in[0,\infty)$.

Because $n(1-\rho)=h_n\to h_1<\infty$, it follows that $h_n=O(1)$ and $\rho_n\to 1$. Hence, $\exp(-h_n^*/n)=\rho_n\to 1$ and $h_n^*=o(n)$. By a mean-value expansion of $\exp(-h_n^*/n)$ about 0,

$$0=\rho_n-\rho_n=\exp(-h_n^*/n)-(1-h_n/n)=h_n/n-\exp(-h_n^{**}/n)h_n^*/n, \tag{10.86}$$

where $h_n^{**}=o(n)$ given that $h_n^*=o(n)$. Hence, $h_n-(1+o(1))h_n^*=0$, and thus $h_n^*/h_n\to 1$. Hence, $h_n^*=O(1)$ and $t_n=bh_n^*/n\to 0$. The proof for $\xi_{1,k}=o_p(1)$ and $\xi_{2,k}=o_p(1)$ used in Case (iii) then goes through.

10.2.5 Verification of Assumption EE

In this section, we verify Assumption EE for model 1. We verify Assumption EE using the same argument as for Assumption E given above, but with $T_{n,k}=S_{1,k}S_{2,k}^{-1/2}$ replaced by $d_{b_n}(\gamma_{n,h})\widehat{\sigma}_{n,b_n,k}$, where $d_{b_n}(\gamma_{n,h})$ is the normalization constant that appears in Assumption BB and is defined in (10.2). In Case (i) of the verification of Assumption E above, where $b(1-\rho)\to\infty$, we have $d_{b_n}(\gamma_{n,h})\widehat{\sigma}_{n,b_n,k}\to_p 1$ by Proposition S1(b). Thus, (10.19) trivially holds in this case. In Cases (ii)-(iv), we have $d_{b_n}(\gamma_{n,h})\widehat{\sigma}_{n,b_n,k}=S_{2,k}^{1/2}S_{3,k}^{-1}$ for $S_{3,k}=b^{-2}X_1'M_{X_2}X_1$, where as above (with abuse of notation) X_1 and X_2 denote b-vectors containing data from the k-th subsample. It is sufficient to show the equivalent of (10.39) for $S_{3,k}$:

$$\begin{aligned}&S_{3,k}=\widetilde{S}_{3,k}+\xi_{3,k}\text{ for some }\widetilde{S}_{3,k}\text{ that is separated from}\\&S_{3,0}\text{ by }b\text{ time periods }\forall k\geq k_n\text{and }\xi_{3,k}=o_p(1).\end{aligned} \tag{10.87}$$

Easy calculations show that $S_{3,k}=b^{-2}\sum_{i=1}^{b}f_{k,i}^2/\phi_{k+i}^2$. Set $\widetilde{S}_{3,k}=b^{-2}\sum_{i=1}^{b}(f_{k,i}^t)^2/\phi_{k+i}^2$ and $\xi_{3,k}=b^{-2}\sum_{i=1}^{b}(f_{k,i}^2-(f_{k,i}^t)^2)/\phi_{k+i}^2$. Then, proceeding exactly as in the verification of $S_{2,k}=\widetilde{S}_{2,k}+\xi_{2,k}$ in (10.39) in the proof of Assumption E, (10.87) follows.

10.2.6 Verification of Assumption HH

Given the definitions in (10.2), Assumption HH holds by the following calculations. For all sequences $\{\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3}) \in \Gamma : n \geq 1\}$ for which $b_n\gamma_{n,h,1} \to g_1$ for some $g_1 \in R_{+,\infty}$, if $b_n\gamma_{n,h,1} \to g_1 = \infty$, then $n\gamma_{n,h,1} \to \infty$ and

$$\frac{a_{b_n}(\gamma_{n,h})}{a_n(\gamma_{n,h})} = \frac{b_n^{1/2} d_{b_n}(\gamma_{n,h})}{n^{1/2} d_n(\gamma_{n,h})} = \left(\frac{b_n}{n}\right)^{1/2} \to 0 \tag{10.88}$$

using Assumption C(ii). If $n\gamma_{n,h,1} \to h_1 = \infty$ and $b_n\gamma_{n,h,1} \to g_1 < \infty$, let $h_{n,1} = n(1-\rho)$ and let $h_{n,1}^*$ be defined by $\rho = \exp(-h_{n,1}^*/n)$. By Lemma S1 and (10.3), $b_n/(n^{1/2}d_n(\gamma_{n,h})) = O((1-\rho)^{1/2}b_n/n^{1/2}) = O((h_{n,1}/n)^{1/2}b_n/n^{1/2}) = O(h_{n,1}^{1/2}b_n/n)$. Given that $n\gamma_{n,h,1} \to h_1 = \infty$ and $b_n\gamma_{n,h,1} \to g_1 < \infty$ we are either in Case (ii) or Case (iii) of the proof of Assumption E. In Case (iii), we showed above that $t_n = b_n h_{n,1}^*/n \to 0$ and $h_{n,1}/h_{n,1}^* \to 1$, and $h_{n,1} \to \infty$. Therefore, $O(h_{n,1}^{1/2}b_n/n) = O(t_n h_{n,1}^{-1/2}) = o(1)$. In Case (ii), $t_n = (b_n h_{n,1}/n)(h_{n,1}^*/h_{n,1}) \to g_1$ and thus $O(h_{n,1}^{1/2}b_n/n) = O(t_n h_{n,1}^{-1/2}) = o(1)$ because $h_{n,1} \to \infty$. Therefore,

$$\frac{a_{b_n}(\gamma_{n,h})}{a_n(\gamma_{n,h})} = \frac{b_n}{n^{1/2}d_n(\gamma_{n,h})} \to 0. \tag{10.89}$$

If $n\gamma_{n,h,1} \to h_1 < \infty$, then

$$\frac{a_{b_n}(\gamma_{n,h})}{a_n(\gamma_{n,h})} = \frac{b_n}{n} \to 0 \tag{10.90}$$

using Assumption C(ii).

11 Conservative Model Selection Example

11.1 The Model

Here we establish the asymptotic distribution of the test statistic $T_n^*(\theta_0)$ and verify Assumption G for this example.

The model is

$$\begin{aligned} y_i &= x_{1i}^*\theta + x_{2i}^*\beta_2 + x_{3i}^{*\prime}\beta_3 + \sigma\varepsilon_i \text{ for } i = 1, ..., n, \text{ where} \\ x_i^* &= (x_{1i}^*, x_{2i}^*, x_{3i}^{*\prime})' \in R^k,\ \beta = (\theta, \beta_2, \beta_3')' \in R^k, \end{aligned} \tag{11.1}$$

$x_{1i}^*, x_{2i}^*, \theta, \beta_2, \sigma, \varepsilon_i \in R$, and $x_{3i}^*, \beta_3 \in R^{k-2}$. The observations $\{(y_i, x_i^*) : i = 1, ..., n\}$ are i.i.d. The scaled error ε_i has mean 0 and variance 1 conditional on x_i^*. We consider testing $H_0 : \theta = \theta_0$ after carrying out a model selection procedure to determine whether x_{2i}^* should enter the model. The model selection procedure is based on a t test of $H_0^* : \beta_2 = 0$.

The inference problem described above covers the following (seemingly more general) inference problem. Consider the model

$$\begin{aligned} y_i &= z_i'\tau + \sigma\varepsilon_i \text{ for } i = 1, ..., n, \text{ where} \\ z_i &= (z_{1i}', z_{2i})' \in R^k,\ \tau = (\tau_1', \tau_2)' \in R^k, \end{aligned} \tag{11.2}$$

$z_{1i}, \tau_1 \in R^{k-1}$, and $z_{2i}, \tau_2 \in R$. We are interested in testing $\overline{H}_0 : a'\tau = \theta_0$ for a given vector $a \in R^k$ with $a \neq e_k$, where $e_k = (0, ..., 0, 1)'$, after using a (fixed critical value) t test to determine whether z_{2i} should enter the model. This testing problem can be transformed into the former one by writing

$$\theta = a'\tau, \beta_2 = \tau_2,\ \beta_3 = B'\tau, \tag{11.3}$$

for some matrix $B \in R^{k\times(k-2)}$ such that $D = [a : e_k : B] \in R^{k\times k}$ is nonsingular. As defined, $\beta = D'\tau$. Define $x_i^* = D^{-1}z_i$. Then, $x_i^{*\prime}\beta = z_i'\tau$ and $H_0 : \theta = \theta_0$ is equivalent to $\overline{H}_0 : a'\tau = \theta_0$.

We now return to the model in (11.1). To define the test statistic $T_n^*(\theta_0)$, we write the variables in matrix notation and define the first and second regressors after projecting out the remaining regressors using finite-sample projections:

$$\begin{aligned} Y &= (y_1, ..., y_n)', \\ X_j^* &= (x_{j1}^*, ..., x_{jn}^*)' \in R^n \text{ for } j = 1, 2, \\ X_3^* &= [x_{31}^* : ... : x_{3n}^*]' \in R^{n\times(k-2)}, \\ X_j &= M_{X_3^*}X_j^* \in R^n \text{ for } j = 1, 2, \text{ and} \\ X &= [X_1 : X_2] \in R^{n\times 2}, \end{aligned} \tag{11.4}$$

where $M_{X_3^*} = I_n - P_{X_3^*}$ and $P_{X_3^*} = X_3^*(X_3^{*\prime}X_3^*)^{-1}X_3^{*\prime}$. The n-vectors X_1 and X_2 correspond to the n-vectors X_1^* and X_2^*, respectively, with X_3^* projected out.

The restricted and unrestricted least squares (LS) estimators of θ and the unrestricted LS estimator of β_2 are

$$\begin{aligned} \widetilde{\theta} &= (X_1'X_1)^{-1}X_1'Y, \\ \widehat{\theta} &= (X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}Y, \text{ and} \\ \widehat{\beta}_2 &= (X_2'M_{X_1}X_2)^{-1}X_2'M_{X_1}Y. \end{aligned} \tag{11.5}$$

The model selection test rejects $H_0^* : \beta_2 = 0$ if

$$\begin{aligned} |T_{n,2}| &= \left| \frac{n^{1/2}\widehat{\beta}_2}{\widehat{\sigma}(n^{-1}X_2'M_{X_1}X_2)^{-1/2}} \right| > c, \text{ where} \\ \widehat{\sigma}^2 &= (n-k)^{-1}Y'M_{[X_1^*:X_2^*:X_3^*]}Y \end{aligned} \tag{11.6}$$

and $c > 0$ is a given critical value that does not depend on n. Typically, $c = z_{1-\alpha/2}$ for some $\alpha > 0$. The estimator $\widehat{\sigma}^2$ of σ^2 is the standard (unrestricted) unbiased estimator.

The test statistic, $T_n^*(\theta_0)$, for testing $H_0 : \theta = \theta_0$ is a t statistic based on the restricted LS estimator of θ when the null hypothesis $H_0^* : \beta_2 = 0$ is not rejected and the unrestricted LS estimator when it is rejected:

$$\begin{aligned} T_n^*(\theta_0) &= \widetilde{T}_{n,1}(\theta_0)1(|T_{n,2}| \leq c) + \widehat{T}_{n,1}(\theta_0)1(|T_{n,2}| > c), \text{ where} \\ \widetilde{T}_{n,1}(\theta_0) &= \frac{n^{1/2}(\widetilde{\theta} - \theta_0)}{\widehat{\sigma}(n^{-1}X_1'X_1)^{-1/2}} \text{ and} \\ \widehat{T}_{n,1}(\theta_0) &= \frac{n^{1/2}(\widehat{\theta} - \theta_0)}{\widehat{\sigma}(n^{-1}X_1'M_{X_2}X_1)^{-1/2}}. \end{aligned} \tag{11.7}$$

Note that both $\widetilde{T}_{n,1}(\theta_0)$ and $\widehat{T}_{n,1}(\theta_0)$ are defined using the unrestricted estimator $\widehat{\sigma}$ of σ. One could define $\widetilde{T}_{n,1}(\theta_0)$ using the restricted LS estimator of σ, but this is not desirable because it leads to an inconsistent estimator of σ under sequences of parameters $\{\beta_2 = \beta_{2n} : n \geq 1\}$ that satisfy $\beta_{2n} \to 0$ and $n^{1/2}\beta_{2n} \nrightarrow 0$ as $n \to \infty$. For subsampling tests, one could define $\widetilde{T}_{n,1}(\theta_0)$ and $\widehat{T}_{n,1}(\theta_0)$ with $\widehat{\sigma}$ deleted because the scale of the subsample statistics offsets that of the original sample statistic. This does not work for hybrid tests because Assumption K fails if $\widehat{\sigma}$ is deleted.

The "model-selection" estimator, $\overline{\theta}$, of θ is

$$\overline{\theta} = \widetilde{\theta}1(|T_{n,2}| \leq c) + \widehat{\theta}1(|T_{n,2}| > c). \tag{11.8}$$

This estimator is used to recenter the subsample statistics. (One could also use the unrestricted estimator $\widehat{\theta}$ to recenter the subsample statistics.)

11.2 Proof of the Asymptotic Distributions of the Test Statistics

In this section, we establish the asymptotic distribution J_h^* of $T_n^*(\theta_0)$ under a sequence of parameters $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) : n \geq 1\}$ (where $n^{1/2}\gamma_{n,1} \to h_1$, $\gamma_{n,2} \to h_2$, and $\gamma_{n,3} \in \Gamma_3(\gamma_{n,1}, \gamma_{n,2})$ for all n). Parts of the proof are closely related to calculations in Leeb (2006) and Leeb and Pötscher (2005). No papers in the literature, that we are aware of, consider subsampling methods for post-model selection inference. For FCV tests, the main differences from Leeb (2006) are that here we consider (i) model selection among two models, (ii) errors that may be non-normal, (iii) i.i.d. regressors, (iv) t statistics, and (v) we prove the asymptotic results directly. In contrast, Leeb (2006) considers (i) multiple models, (ii) normal errors, (iii) fixed regressors, (iv) normalized estimators, and (v) he proves the asymptotic results by establishing finite-sample results for the normal error case and taking their limits. The results in Leeb and Pötscher (2005) are a two-model special case of those given in Leeb (2006).

Using the definition of $T_n^*(\theta_0)$ in this example, we have

$$\begin{aligned} P_{\theta_0,\gamma_n}(T_n^*(\theta_0) \leq x) = \; & P_{\theta_0,\gamma_n}(\widetilde{T}_{n,1}(\theta_0) \leq x \;\&\; |T_{n,2}| \leq c) \\ & + P_{\theta_0,\gamma_n}(\widehat{T}_{n,1}(\theta_0) \leq x \;\&\; |T_{n,2}| > c). \end{aligned} \tag{11.9}$$

Hence, it suffices to determine the limits of the two summands on the right-hand side. With this in mind, we show below that under $\{\gamma_n : n \geq 1\}$, when $|h_1| < \infty$,

$$\begin{pmatrix} \widetilde{T}_{n,1}(\theta_0) \\ T_{n,2} \end{pmatrix} \to_d \begin{pmatrix} \widetilde{Z}_{h,1} \\ Z_{h,2} \end{pmatrix} \sim N\left(\begin{pmatrix} -h_1 h_2 (1-h_2^2)^{-1/2} \\ h_1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \text{ and}$$
$$\begin{pmatrix} \widehat{T}_{n,1}(\theta_0) \\ T_{n,2} \end{pmatrix} \to_d \begin{pmatrix} \widehat{Z}_{h,1} \\ Z_{h,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ h_1 \end{pmatrix}, \begin{pmatrix} 1 & h_2 \\ h_2 & 1 \end{pmatrix}\right). \quad (11.10)$$

Given this, we have

$$\begin{aligned} & P_{\theta_0,\gamma_n}(\widetilde{T}_{n,1}(\theta_0) \leq x \ \& \ |T_{n,2}| \leq c) \\ \to\ & P(\widetilde{Z}_{h,1} \leq x \ \& \ |Z_{h,2}| \leq c) \\ =\ & \Phi(x + h_1 h_2 (1-h_2^2)^{-1/2}) \Delta(h_1, c), \text{ where} \\ & \Delta(a,b) = \Phi(a+b) - \Phi(a-b), \end{aligned} \quad (11.11)$$

the equality uses the independence of $\widetilde{Z}_{h,1}$ and $Z_{h,2}$ and the normality of their distributions, and $\Delta(a,b) = \Delta(-a,b)$. In addition, we have

$$P_{\theta_0,\gamma_n}(\widehat{T}_{n,1}(\theta_0) \leq x \ \& \ |T_{n,2}| > c) \to P(\widehat{Z}_{h,1} \leq x \ \& \ |Z_{h,2}| > c). \quad (11.12)$$

Next, we calculate the limiting probability in (11.12). Let $f(z_2|z_1)$ denote the conditional density of $Z_{h,2}$ given $\widehat{Z}_{h,1}$. Let $\phi(z_1)$ denote the standard normal density. Given that

$$\begin{pmatrix} \widehat{Z}_{h,1} \\ Z_{h,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ h_1 \end{pmatrix}, \begin{pmatrix} 1 & h_2 \\ h_2 & 1 \end{pmatrix}\right), \quad (11.13)$$

the conditional distribution of $Z_{h,2}$ given $\widehat{Z}_{h,1} = z_1$ is $N(h_1 + h_2 z_1, 1 - h_2^2)$. We have

$$\begin{aligned} & P(\widehat{Z}_{h,1} \leq x \ \& \ |Z_{h,2}| > c) \\ =\ & \int_{-\infty}^{x} \int_{|z_2|>c} f(z_2|z_1)\phi(z_1) dz_2 dz_1 \\ =\ & \int_{-\infty}^{x} \left(1 - \int_{|z_2| \leq c} (1-h_2^2)^{-1/2} \phi\left(\frac{z_2 - (h_1 + h_2 z_1)}{(1-h_2^2)^{1/2}}\right) dz_2\right) \phi(z_1) dz_1 \\ =\ & \int_{-\infty}^{x} \left(1 - \int_{|\overline{z}_2| \leq c(1-h_2^2)^{-1/2}} \phi\left(\overline{z}_2 - \frac{h_1 + h_2 z_1}{(1-h_2^2)^{1/2}}\right) d\overline{z}_2\right) \phi(z_1) dz_1 \\ =\ & \int_{-\infty}^{x} \left(1 - \Delta\left(\frac{h_1 + h_2 z}{(1-h_2^2)^{1/2}}, \frac{c}{(1-h_2^2)^{1/2}}\right)\right) \phi(z) dz, \end{aligned} \quad (11.14)$$

where the second equality holds by (11.13), the third equality holds by change of variables with $\overline{z}_2 = z_2(1-h_2^2)^{-1/2}$, and the last equality holds by the definition of $\Delta(a,b)$.

Combining (11.11), (11.12), and (11.14) gives the desired result:

$$J_h^*(x) = \Phi(x + h_1 h_2(1-h_2^2)^{-1/2})\Delta(h_1, c) + \int_{-\infty}^{x} \left(1 - \Delta\left(\frac{h_1 + h_2 t}{(1-h_2^2)^{1/2}}, \frac{c}{(1-h_2^2)^{1/2}}\right)\right)\phi(t)dt \tag{11.15}$$

when $|h_1| < \infty$. When $|h_1| = \infty$, $J_h^*(x) = \Phi(x)$ (which equals the limit as $|h_1| \to \infty$ of $J_h^*(x)$ defined in (11.15)). The proof of the latter result is given below in the paragraph containing (11.29).

We now show that under $\{\gamma_n : n \geq 1\}$, when $|h_1| < \infty$, (11.10) holds. Let $X_j^\perp = (x_{j1}^\perp, ..., x_{jn}^\perp)' \in R^n$ for $j = 1, 2$ and $X^\perp = (X_1^\perp, X_2^\perp)' \in R^{n\times 2}$. We use the following Lemma.

Lemma S3. *Given the assumptions stated in* Section 2.2 *of the paper, under a sequence of parameters* $\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) : n \geq 1\}$ *(where* $n^{1/2}\gamma_{n,1} \to h_1$, $\gamma_{n,2} \to h_2$, *and* $\gamma_{n,3} \in \Gamma_3(\gamma_1, \gamma_2)$ *for all* n*), and for* $Q = Q_n$ *as defined in* (2.7) *of the paper with the* (j, m) *element denoted* $Q_{n,jm}$*, we have*
(a) $n^{-1}X'X - Q_n \to_p 0$, (b) $n^{-1}X_2'M_{X_1}X_2 - (Q_{n,22} - Q_{n,12}^2 Q_{n,11}^{-1}) \to_p 0$, (c) $n^{-1}X_1'M_{X_2}X_1 - (Q_{n,11} - Q_{n,12}^2 Q_{n,22}^{-1}) \to_p 0$, (d) $\widehat{\sigma}/\sigma_n \to_p 1$, (e) $n^{-1/2}X_j'\varepsilon = n^{-1/2}X_j^{\perp\prime}\varepsilon + o_p(1) = O_p(1)$ *for* $j = 1, 2$.

Proof of Lemma S3. The proofs of parts (a)-(d) are standard using a weak law of large numbers (WLLN) for $L^{1+\delta}$-bounded independent random variables for some $\delta > 0$ and taking into account the fact that $X_j = M_{X_3^*}X_j^*$ for $j = 1, 2$.

Next, we prove part (e). By definition of X_j, we have

$$\begin{aligned} n^{-1/2}X_j'\varepsilon &= n^{-1/2}X_j^{*\prime}\varepsilon - n^{-1}X_j^{*\prime}X_3^*(n^{-1}X_3^{*\prime}X_3^*)^{-1}n^{-1/2}X_3^{*\prime}\varepsilon \\ &= n^{-1/2}X_j^{*\prime}\varepsilon - E_{G_n}x_{ji}^*x_{3i}^{*\prime}(E_{G_n}x_{3i}^*x_{3i}^{*\prime})^{-1}n^{-1/2}X_3^{*\prime}\varepsilon + o_p(1) \\ &= n^{-1/2}X_j^{\perp\prime}\varepsilon + o_p(1), \end{aligned} \tag{11.16}$$

where G_n denotes the distribution of (ε_i, x_i^*) under γ_n, the second equality holds by the same WLLN as above combined with the Lindeberg triangular array central limit theorem (CLT) applied to $n^{-1/2}X_3^{*\prime}\varepsilon$, which yields $n^{-1/2}X_3^{*\prime}\varepsilon = O_p(1)$, and the third equality uses the definition that $x_{ji}^\perp = x_{ji}^* - E_{G_n}x_{ji}^*x_{3i}^{*\prime}(E_{G_n}x_{3i}^*x_{3i}^{*\prime})^{-1}x_{3i}^*$. The second equality of part (e) holds by the Lindeberg CLT. The Lindeberg condition is implied by a Liapounov condition, which holds by the moment bound in $\Gamma_3(\gamma_1, \gamma_2)$. $\square$

We now prove the first result of (11.10) (which assumes $|h_1| < \infty$). Using (11.5) and (11.6), we have

$$T_{n,2} = \frac{n^{1/2}\beta_2/\sigma_n + (n^{-1}X_2'M_{X_1}X_2)^{-1}n^{-1/2}X_2'M_{X_1}\varepsilon}{(\widehat{\sigma}/\sigma_n)(n^{-1}X_2'M_{X_1}X_2)^{-1/2}}$$

$$= n^{1/2}\frac{\beta_2}{\sigma_n(Q_n^{22})^{1/2}}(1+o_p(1)) + (Q_n^{22})^{1/2}n^{-1/2}X_2'(I_n - P_{X_1})\varepsilon(1+o_p(1))$$
$$= n^{1/2}\gamma_{n,1}(1+o_p(1)) + (Q_n^{22})^{1/2}(e_2 - Q_{n,12}Q_{n,11}^{-1}e_1)'n^{-1/2}X'\varepsilon(1+o_p(1)),$$
$$= h_1 + (Q_n^{22})^{1/2}(e_2 - Q_{n,12}Q_{n,11}^{-1}e_1)'n^{-1/2}X^{\perp\prime}\varepsilon + o_p(1), \quad (11.17)$$

where $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$, $e_1 = (1,0)'$, $e_2 = (0,1)'$, $Q_n = E_{G_n}x_i^{\perp}x_i^{\perp\prime}$, Q_n^{22} is the $(2,2)$ element of Q_n^{-1}, the second equality uses Lemma S3(b) and (d), the fact that $Q_n^{22} = (Q_{n,22} - Q_{n,12}^2Q_{n,11}^{-1})^{-1}$, and the fact that $\lambda_{\min}(Q_n) \geq \kappa > 0$ by definition of $\Gamma_3(\gamma_1, \gamma_2)$, the third equality uses the definition of $\gamma_{n,1}$ and Lemma S3(a), and the fourth equality holds by the assumption that $n^{1/2}\gamma_{n,1} \to h_1$ and Lemma S3(e).

Using (11.5) and (11.7), we have

$$\widetilde{T}_{n,1}(\theta_0) = \frac{n^{1/2}(n^{-1}X_1'X_1)^{-1}n^{-1}X_1'X_2\beta_2/\sigma_n + (n^{-1}X_1'X_1)^{-1}n^{-1/2}X_1'\varepsilon}{(\widehat{\sigma}/\sigma_n)(n^{-1}X_1'X_1)^{-1/2}}$$
$$= n^{1/2}\frac{Q_{n,12}\beta_2}{\sigma_n Q_{n,11}^{1/2}}(1+o_p(1)) + Q_{n,11}^{-1/2}n^{-1/2}e_1'X'\varepsilon(1+o_p(1))$$
$$= h_1\frac{Q_{n,12}(Q_n^{22})^{1/2}}{Q_{n,11}^{1/2}} + Q_{n,11}^{-1/2}n^{-1/2}e_1'X^{\perp\prime}\varepsilon + o_p(1), \quad (11.18)$$

where the second equality uses Lemma S3(a) and (d), and the third equality uses the assumption that $n^{1/2}\gamma_{n,1} = n^{1/2}\beta_2/(\sigma_n^2Q_n^{22})^{1/2} \to h_1$ and Lemma S3(e).

We have

$$Q_n^{-1} = \frac{1}{Q_{n,11}Q_{n,22} - Q_{n,12}^2}\begin{bmatrix} Q_{n,22} & -Q_{n,12} \\ -Q_{n,12} & Q_{n,11}\end{bmatrix} \text{ and so}$$
$$\gamma_{n,2} = \frac{Q_n^{12}}{(Q_n^{11}Q_n^{22})^{1/2}} = \frac{-Q_{n,12}}{(Q_{n,11}Q_{n,22})^{1/2}} \text{ and}$$
$$Q_n^{22} = \frac{Q_{n,11}}{Q_{n,11}Q_{n,22} - Q_{n,12}^2} = (Q_{n,22})^{-1}(1-\gamma_{n,2}^2)^{-1}, \quad (11.19)$$

where the first equality in the second line holds by the definition of $\gamma_{n,2}$ in (2.6) of the paper. This yields

$$\frac{Q_{n,12}(Q_n^{22})^{1/2}}{Q_{n,11}^{1/2}} = \frac{Q_{n,12}(1-\gamma_{n,2}^2)^{-1/2}}{Q_{n,11}^{1/2}Q_{n,22}^{1/2}} = -\gamma_{n,2}(1-\gamma_{n,2}^2)^{-1/2} = -h_2(1-h_2^2)^{-1/2} + o(1). \quad (11.20)$$

Combining (11.17), (11.18), and (11.20) gives

$$\begin{pmatrix} \widetilde{T}_{n,1}(\theta_0) \\ T_{n,2} \end{pmatrix} = \begin{pmatrix} -h_1h_2(1-h_2^2)^{-1/2} + Q_{n,11}^{-1/2}n^{-1/2}e_1'X^{\perp\prime}\varepsilon \\ h_1 + (Q_n^{22})^{1/2}(e_2 - Q_{n,12}Q_{n,11}^{-1}e_1)'n^{-1/2}X^{\perp\prime}\varepsilon \end{pmatrix} + o_p(1). \quad (11.21)$$

The first result of (11.10) holds by (11.21), the Lindeberg CLT, and the Cramér-Wold device. The Lindeberg condition is implied by a Liapounov condition, which holds by the moment bound in $\Gamma_3(\gamma_1, \gamma_2)$. The asymptotic covariance matrix is I_2 by the following calculations. The $(1, 2)$ element of the asymptotic covariance matrix equals

$$\begin{aligned} & E_{G_n} Q_{n,11}^{-1/2} e_1' n^{-1} X^{\perp\prime} X^{\perp} (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)(Q_n^{22})^{1/2} \\ = \; & Q_{n,11}^{-1/2} e_1' Q_n (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)(Q_n^{22})^{1/2} = 0, \end{aligned} \tag{11.22}$$

where the first equality holds because $E_{G_n} x_i^{\perp} x_i^{\perp\prime} = Q_n$ and the second equality holds by algebra. The $(1, 1)$ element equals

$$E_{G_n} Q_{n,11}^{-1/2} e_1' n^{-1} X^{\perp\prime} X^{\perp} e_1 Q_{n,11}^{-1/2} = Q_{n,11}^{-1/2} e_1' Q_n e_1 Q_{n,11}^{-1/2} = 1. \tag{11.23}$$

The $(2, 2)$ element equals

$$\begin{aligned} & E_{G_n} (Q_n^{22})^{1/2} (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)' n^{-1} X^{\perp\prime} X^{\perp} (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)(Q_n^{22})^{1/2} \\ = \; & (Q_n^{22})^{1/2} (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)' Q_n (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)(Q_n^{22})^{1/2} \\ = \; & (Q_n^{22})^{1/2} (Q_{n,22}(1 - \gamma_{n,2}^2))(Q_n^{22})^{1/2} = 1, \end{aligned} \tag{11.24}$$

where the second equality holds by algebra and the definition of $\gamma_{n,2}$ and the third equality holds by the third result in (11.19). This completes the proof of the first result in (11.10).

Next, we prove the second result in (11.10). Using (11.7), we have

$$\begin{aligned} \widehat{T}_{n,1}(\theta_0) & = \frac{(n^{-1} X_1' M_{X_2} X_1)^{-1} n^{-1/2} X_1' M_{X_2} \varepsilon}{(\widehat{\sigma}/\sigma_n)(n^{-1} X_1' M_{X_2} X_1)^{-1/2}} \\ & = (Q_n^{11})^{1/2} (e_1 - Q_{n,12} Q_{n,22}^{-1} e_2)' n^{-1/2} X^{\perp\prime} \varepsilon + o_p(1), \end{aligned} \tag{11.25}$$

where the second equality holds analogously to (11.17). Combining (11.17) and (11.25) gives

$$\begin{pmatrix} \widehat{T}_{n,1}(\theta_0) \\ T_{n,2} \end{pmatrix} = \begin{pmatrix} (Q_n^{11})^{1/2} (e_1 - Q_{n,12} Q_{n,22}^{-1} e_2)' n^{-1/2} X^{\perp\prime} \varepsilon \\ h_1 + (Q_n^{22})^{1/2} (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)' n^{-1/2} X^{\perp\prime} \varepsilon \end{pmatrix} + o_p(1). \tag{11.26}$$

The second result of (11.10) holds by (11.26), the Lindeberg CLT, and the Cramér-Wold device. The Lindeberg condition holds as above. The 2×2 asymptotic covariance matrix has off-diagonal element h_2 and diagonal elements equal to one by the following calculations. The $(1, 2)$ element equals

$$\begin{aligned} & E_{G_n} (Q_n^{11})^{1/2} (e_1 - Q_{n,12} Q_{n,22}^{-1} e_2)' n^{-1} X^{\perp\prime} X^{\perp} (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)(Q_n^{22})^{1/2} \\ = \; & (Q_n^{11})^{1/2} (e_1 - Q_{n,12} Q_{n,22}^{-1} e_2)' Q_n (e_2 - Q_{n,12} Q_{n,11}^{-1} e_1)(Q_n^{22})^{1/2} \\ = \; & (Q_n^{11})^{1/2} (-Q_{n,12}(1 - Q_{n,12}^2 Q_{n,11}^{-1} Q_{n,22}^{-1}))(Q_n^{22})^{1/2} \\ = \; & (Q_{n,11}(1 - \gamma_{n,2}^2))^{-1/2} (-Q_{n,12}(1 - \gamma_{n,2}^2))(Q_{n,22}(1 - \gamma_{n,2}^2))^{-1/2} \\ = \; & \frac{-Q_{n,12}}{(Q_{n,11} Q_{n,22})^{1/2}} = \frac{Q_n^{12}}{(Q_n^{11} Q_n^{22})^{1/2}} = \gamma_{n,2} = h_2 + o(1), \end{aligned} \tag{11.27}$$

where the second equality holds by algebra, the third equality holds by the second and third results of (11.19) and the third result of (11.19) with 22 and 11 interchanged, and the fifth and sixth equalities hold by the second result of (11.19).

The $(1,1)$ element equals

$$\begin{aligned} & E_{G_n}(Q_n^{11})^{1/2}(e_1 - Q_{n,12}Q_{n,22}^{-1}e_2)'n^{-1}X^{\perp\prime}X^{\perp}(e_1 - Q_{n,12}Q_{n,22}^{-1}e_2)(Q_n^{11})^{1/2} \\ = \; & (Q_n^{11})^{1/2}(e_1 - Q_{n,12}Q_{n,22}^{-1}e_2)'Q_n(e_1 - Q_{n,12}Q_{n,22}^{-1}e_2)(Q_n^{11})^{1/2} = 1, \end{aligned} \quad (11.28)$$

where the second equality holds by an analogous argument to that in (11.24). The $(2,2)$ element equals one by (11.24). This completes the proof of the second result in (11.10).

Finally, we show that $J_h^*(x) = \Phi(x)$ when $|h_1| = \infty$. Equations (11.25) and (11.28) hold in this case, so $\widehat{T}_{n,1}(\theta_0) \to_d N(0,1)$ under $\{\gamma_n : n \geq 1\}$. The first three equalities of (11.17) hold when $|h_1| = \infty$ and show that $|T_{n,2}| \to_p \infty$. These results combine to yield

$$\begin{aligned} & P_{\theta_0,\gamma_n}(\widetilde{T}_{n,1}(\theta_0) \leq x \;\&\; |T_{n,2}| \leq c) = o(1) \text{ and} \quad (11.29) \\ & P_{\theta_0,\gamma_n}(\widehat{T}_{n,1}(\theta_0) \leq x \;\&\; |T_{n,2}| > c) = P_{\theta_0,\gamma_n}(\widehat{T}_{n,1}(\theta_0) \leq x) + o(1) \to \Phi(x) \end{aligned}$$

for all $x \in R$. This and (11.9) combine to give $P_{\theta_0,\gamma_n}(T_n^*(\theta_0) \leq x) \to \Phi(x)$ and $J_h^*(x) = \Phi(x)$ when $|h_1| = \infty$.

11.3 Verification of Assumption G

Assumption G is verified in the conservative model selection example by using a variant of the argument in the proof of Lemma 4 in AG1 with $\tau_n = a_n = n^{1/2}$ and $d_n = 1$. In the present case, (8.16) of AG1 holds with

$$\begin{aligned} R_n(t) &= q_n^{-1}\sum_{j=1}^{q_n} 1(|b_n^{1/2}(\overline{\theta} - \theta_0)/\widehat{\sigma}_{n,b,j}^{(1)}| \geq t) \\ &\quad + q_n^{-1}\sum_{j=1}^{q_n} 1(|b_n^{1/2}(\overline{\theta} - \theta_0)/\widehat{\sigma}_{n,b,j}^{(2)}| \geq t), \text{ where} \\ \widehat{\sigma}_{n,b,j}^{(1)} &= \widehat{\sigma}_{n,b,j}(b_n^{-1}X_{1,n,b,j}'X_{1,n,b,j})^{-1/2}, \\ \widehat{\sigma}_{n,b,j}^{(2)} &= \widehat{\sigma}_{n,b,j}(b_n^{-1}X_{1,n,b,j}'M_{X_{2,n,b,j}}X_{1,n,b,j})^{-1/2}, \end{aligned} \quad (11.30)$$

and $(X_{1,n,b,j},\ X_{2,n,b,j}, \widehat{\sigma}_{n,b,j})$ denotes $(X_1, X_2, \widehat{\sigma})$ based on the jth subsample rather than the full sample. (Equation (8.16) of AG1 holds with $R_n(t)$ defined as in (11.30) for all three versions of the tests: $T_n(\theta_0) = T_n^*(\theta_0)$, $-T_n^*(\theta_0)$, and $|T_n^*(\theta_0)|$.) As in the proof of Lemma 4 of AG1, it suffices to show that $R_n(t)$ converges in probability to zero under all sequences $\{\gamma_{n,h} : n \geq 1\}$ for all $t > 0$. The assumption that $b_n/n \to 0$

and the result established below that $n^{1/2}(\overline{\theta} - \theta_0)/\sigma_n = O_p(1)$ under all sequences $\{\gamma_{n,h} : n \geq 1\}$ imply that for all $\delta > 0$, wp$\rightarrow 1$,

$$R_n(t) \leq R_n^{(1)}(\delta, t) + R_n^{(2)}(\delta, t), \text{ where } R_n^{(m)}(\delta, t) = q_n^{-1} \sum_{j=1}^{q_n} 1(\delta\sigma_n/\widehat{\sigma}_{n,b,j}^{(m)} \geq t) \quad (11.31)$$

for $m = 1, 2$. The variance of $R_n^{(m)}(\delta, t)$ goes to zero under $\{\gamma_{n,h} : n \geq 1\}$ by the same U-statistic argument for i.i.d. observations as used to establish Assumption E of AG1 in the i.i.d. case, see Section 3.3 of AG1. The expectation of $R_n^{(m)}(\delta, t)$ equals $P_{\theta_0,\gamma_{n,h}}(\widehat{\sigma}_{n,b,j}^{(m)}/\sigma_n \leq \delta/t)$. We have

$$\widehat{\sigma}_{n,b,j}^{(1)}/\sigma_n = (\widehat{\sigma}_{n,b,j}/\sigma_n)[(b_n^{-1}X_{1,n,b,j}'X_{1,n,b,j})^{-1/2} - Q_{n,11}^{-1/2} + Q_{n,11}^{-1/2}] = Q_{n,11}^{-1/2} + o_p(1), \quad (11.32)$$

where the second equality holds by Lemma S3 (or, more precisely, by the same argument as used to prove Lemma S3). In addition, $Q_{n,11}^{-1/2}$ is bounded away from zero as $n \rightarrow \infty$ by the definition of $\Gamma_3(\gamma_1, \gamma_2)$. In consequence, the expectation of $R_n^{(1)}(\delta, t)$ goes to zero for all δ sufficiently small. Since the mean and variance of $R_n^{(1)}(\delta, t)$ go to zero, $R_n^{(1)}(\delta, t) \rightarrow_p 0$ for $\delta > 0$ sufficient small. An analogous argument shows that $R_n^{(2)}(\delta, t) \rightarrow_p 0$ for $\delta > 0$ sufficient small. These results and (11.31) yield $R_n(t) \rightarrow_p 0$ under all sequences $\{\gamma_{n,h} : n \geq 1\}$, as desired.

It remains to show that $n^{1/2}(\overline{\theta}-\theta_0)/\sigma_n = O_p(1)$ under all sequences $\{\gamma_{n,h} : n \geq 1\}$. We consider two cases: $|h_1| = \infty$ and $|h_1| < \infty$. First, suppose $|h_1| = \infty$. Then, the first three equalities of (11.17) hold and show that $|T_{n,2}| \rightarrow_p \infty$. In addition, $n^{1/2}(\widehat{\theta} - \theta_0)/\sigma_n = (\widehat{\sigma}/\sigma_n)(n^{-1}X_1'M_{X_2}X_1)^{-1/2}\widehat{T}_{n,1}(\theta_0) = O_p(1)$ by (11.10), Lemma S3(c), and the definition of $\Gamma_3(\gamma_1, \gamma_2)$. Combining these results gives: when $|h_1| = \infty$,

$$\begin{aligned} n^{1/2}(\overline{\theta} - \theta_0)/\sigma_n &= [n^{1/2}(\widetilde{\theta} - \theta_0)/\sigma_n]1(|T_{n,2}| \leq c) + [n^{1/2}(\widehat{\theta} - \theta_0)/\sigma_n]1(|T_{n,2}| > c)) \\ &= o_p(1) + O_p(1). \end{aligned} \quad (11.33)$$

Next, suppose $|h_1| < \infty$, then $\widehat{T}_{n,1}(\theta_0) = O_p(1)$ and $\widetilde{T}_{n,1}(\theta_0) = O_p(1)$ by (11.10). In addition, $\widehat{\sigma}/\sigma_n \rightarrow_p 1$, $(n^{-1}X_1'X_1)^{-1/2} = O_p(1)$, and $(n^{-1}X_1'M_{X_2}X_1)^{-1/2} = O_p(1)$ by Lemma S3 and the definition of $\Gamma_3(\gamma_1, \gamma_2)$. Combining these results gives: when $|h_1| < \infty$,

$$\begin{aligned} n^{1/2}(\overline{\theta} - \theta_0)/\sigma_n &= [n^{1/2}(\widetilde{\theta} - \theta_0)/\sigma_n]1(|T_{n,2}| \leq c) + [n^{1/2}(\widehat{\theta} - \theta_0)/\sigma_n]1(|T_{n,2}| > c)) \\ &= (\widehat{\sigma}/\sigma_n)(n^{-1}X_1'X_1)^{-1/2}\widetilde{T}_{n,1}(\theta_0)1(|T_{n,2}| \leq c) \\ &\quad +(\widehat{\sigma}/\sigma_n)(n^{-1}X_1'M_{X_2}X_1)^{-1/2}\widehat{T}_{n,1}(\theta_0)1(|T_{n,2}| > c)) \\ &= O_p(1), \end{aligned} \quad (11.34)$$

which completes the verification of Assumption G.

References

Andrews, D. W. K. and P. Guggenberger (2005): "Applications of Subsampling, Hybrid, and Size-Correction Methods," Cowles Foundation Discussion Paper No. 1608, Yale University.

——— (2008): "Asymptotics for LS, GLS, and Feasible GLS Statistics in an AR(1) Model with Conditional Heteroskedaticity," unpublished manuscript, Cowles Foundation, Yale University.

——— (2009a): "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap," *Econometric Theory,* 25, forthcoming.

——— (2009b): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory,* 25, forthcoming.

Doukhan P. (1994): *Mixing Properties and Examples.* Lecture Notes in Statistics 85. New York: Springer.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion.* New York: Springer–Verlag.

Kristensen, D. and O. Linton (2006): "A Closed-Form Estimator for the GARCH(1,1) Model," *Econometric Theory,* 22, 323-337.

Leeb, H. (2006): "The Distribution of a Linear Predictor After Model Selection: Unconditional Finite-Sample Distributions and Asymptotic Approximations," *2nd Lehmann Symposium—Optimality.* Institute of Mathematical Statistics Lecture Notes—Monograph Series.

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

White H. (1984): *Asymptotic Theory for Econometricians.* New York: Academic Press.

TABLE B-I

NUISANCE PARAMETER NEAR A BOUNDARY EXAMPLE: MAXIMUM (OVER h_1) NULL REJECTION PROBABILITIES ($\times 100$) FOR DIFFERENT VALUES OF THE CORRELATION h_2 FOR VARIOUS NOMINAL 5% TESTS FOR $n = 120$ AND $b = 12$, WHERE THE PROBABILITIES ARE ASYMPTOTIC, FINITE-SAMPLE-ADJUSTED ASYMPTOTIC, AND EXACT

	Test:	Sub	Sub	Sub	PSC-Sub	APSC-Sub	FCV	FCV	PSC-FCV	Hyb	Hyb	Hyb
h_2	Prob:	Asy	Adj-Asy	Exact	Exact	Exact	Asy	Exact	Exact	Asy	Adj-Asy	Exact
(a) Upper 1-Sided Tests												
-1.0		50.2	49.5	49.8	4.9	13.5	5.0	5.2	5.1	5.0	5.0	5.2
-.95		33.8	22.9	25.6	5.1	9.0	5.0	5.2	5.1	5.0	5.0	5.2
-.80		20.2	12.1	13.1	3.1	6.2	5.0	5.1	4.9	5.0	5.0	4.7
-.40		8.3	6.5	5.9	4.8	4.6	5.0	4.9	4.8	5.0	5.0	3.7
.00		5.0	5.0	5.0	4.9	4.9	5.0	5.2	5.0	5.0	5.0	3.7
.20		5.0	5.0	4.9	5.2	4.9	5.6	5.7	5.2	5.0	5.0	3.8
.40		5.0	5.0	5.0	5.0	5.0	5.8	5.8	5.0	5.0	5.0	3.8
.60		5.0	5.0	5.3	5.3	5.3	5.6	5.7	5.1	5.0	5.0	3.9
.90		5.0	5.0	4.9	4.9	4.9	5.0	5.0	4.9	5.0	5.0	3.4
1.00		5.0	5.0	4.8	4.9	4.8	5.0	5.0	4.9	5.0	5.0	3.5
Max		50.2	49.5	49.8	5.3	13.5	5.8	5.8	5.2	5.0	5.0	5.2
(b) Symmetric 2-Sided Tests												
.00		5.0	5.0	5.2	5.1	5.1	5.0	5.4	-	5.0	5.0	3.5
.20		5.2	5.2	5.2	4.9	5.0	5.0	5.3	-	5.0	5.0	3.5
.40		6.0	5.6	5.4	4.5	4.8	5.0	5.2	-	5.0	5.0	3.5
.60		7.5	6.5	6.0	4.0	4.6	5.0	5.3	-	5.0	5.0	3.7
.80		9.6	8.3	6.9	3.7	4.5	5.0	5.2	-	5.0	5.0	3.9
.95		10.1	10.0	8.3	4.2	4.2	5.0	5.7	-	5.0	5.0	4.5
1.00		10.1	10.1	8.4	4.1	4.1	5.0	5.1	-	5.0	5.0	4.2
Max		10.1	10.1	8.4	5.1	5.1	5.0	5.7	-	5.0	5.0	4.5
(c) Equal-Tailed 2-Sided Tests												
.00		5.0	5.0	5.7	5.5	5.5	5.0	5.4	-	5.0	5.0	3.5
.20		5.4	5.2	5.9	5.4	5.6	5.0	5.3	-	5.0	5.0	3.6
.40		6.7	5.8	6.2	4.5	5.4	5.0	5.2	-	5.0	5.0	3.4
.60		9.9	7.0	7.8	3.9	5.7	5.0	5.3	-	5.0	5.0	3.8
.80		17.3	10.3	12.4	3.2	6.5	5.0	5.2	-	5.0	5.0	4.1
.95		32.4	21.0	24.3	3.5	9.1	5.0	5.7	-	5.0	5.0	4.7
1.00		52.7	51.8	52.7	4.6	13.5	5.0	5.1	-	5.0	5.0	4.2
Max		52.7	51.8	52.7	5.5	13.5	5.0	5.7	-	5.0	5.0	4.7

TABLE B-II

AR EXAMPLE: CI COVERAGE PROBABILITIES (×100) FOR NOMINAL 95% CIs

Case	DGP	n=131 or *Asy*	Upper CIs FCV	Upper CIs Sub	Upper CIs Hyb	Lower CIs FCV	Lower CIs Sub	Lower CIs Hyb
		-.90	90.0	93.9	94.4	95.0	94.1	96.1
		-.50	92.8	92.7	94.7	92.6	92.6	94.4
		.00	93.8	89.9	94.5	91.8	95.1	95.6
(i)	GARCH	ρ =.70	95.9	83.6	95.9	88.4	97.7	97.7
	MA=.15,	.80	96.7	83.2	96.7	86.7	97.8	97.8
	AR=.80	.90	97.7	83.9	97.7	84.0	97.9	97.9
	$h_{27} = .86$	.97	98.9	89.2	98.9	74.7	97.5	97.5
		1.0	99.6	95.5	99.6	53.6	95.1	95.1
		FS-Min	90.0	83.2	93.9	53.6	92.4	94.4
		Asy	95.0	57.3	95.0	63.9	95.0	95.0
		Adj-Asy	-	82.6	95.3	-	95.0	95.2
(ii)	IGARCH							
	MA=.20,	FS-Min	90.7	82.2	93.9	56.2	92.4	94.6
	AR=.80							
(iii)	GARCH	FS-Min	90.4	86.2	94.1	60.6	92.9	95.0
	MA=.70,	Asy	95.0	77.2	95.0	79.2	95.0	95.0
	AR=.20	Adj-Asy	-	88.7	95.3	-	94.8	95.1
	$h_{27} = .54$							
(iv)	i.i.d.	FS-Min	90.5	82.8	94.0	50.4	92.7	94.3
	$h_{27} = 1$	Asy	95.2	47.4	95.0	53.4	95.0	95.0
		Adj-Asy	-	78.8	95.0	-	95.1	95.0
(v)	ARCH4	FS-Min	90.5	84.4	93.9	58.8	92.8	94.8
	(.3,.2,.2,.2)	Asy	95.0	77.2	95.0	79.2	95.0	95.0
	$h_{27} = .54$	Adj-Asy	-	88.7	95.3	-	94.8	95.1
(vi)	IARCH4	FS-Min	90.5	84.0	93.9	60.8	92.4	94.7
	(.3,.3,.2,.2)							
	Min Over	Asy	94.8	47.5	94.8	54.5	94.9	94.9
	$h_{27} \in [0, 1]$	Adj-Asy	-	78.8	95.1	-	94.8	95.0

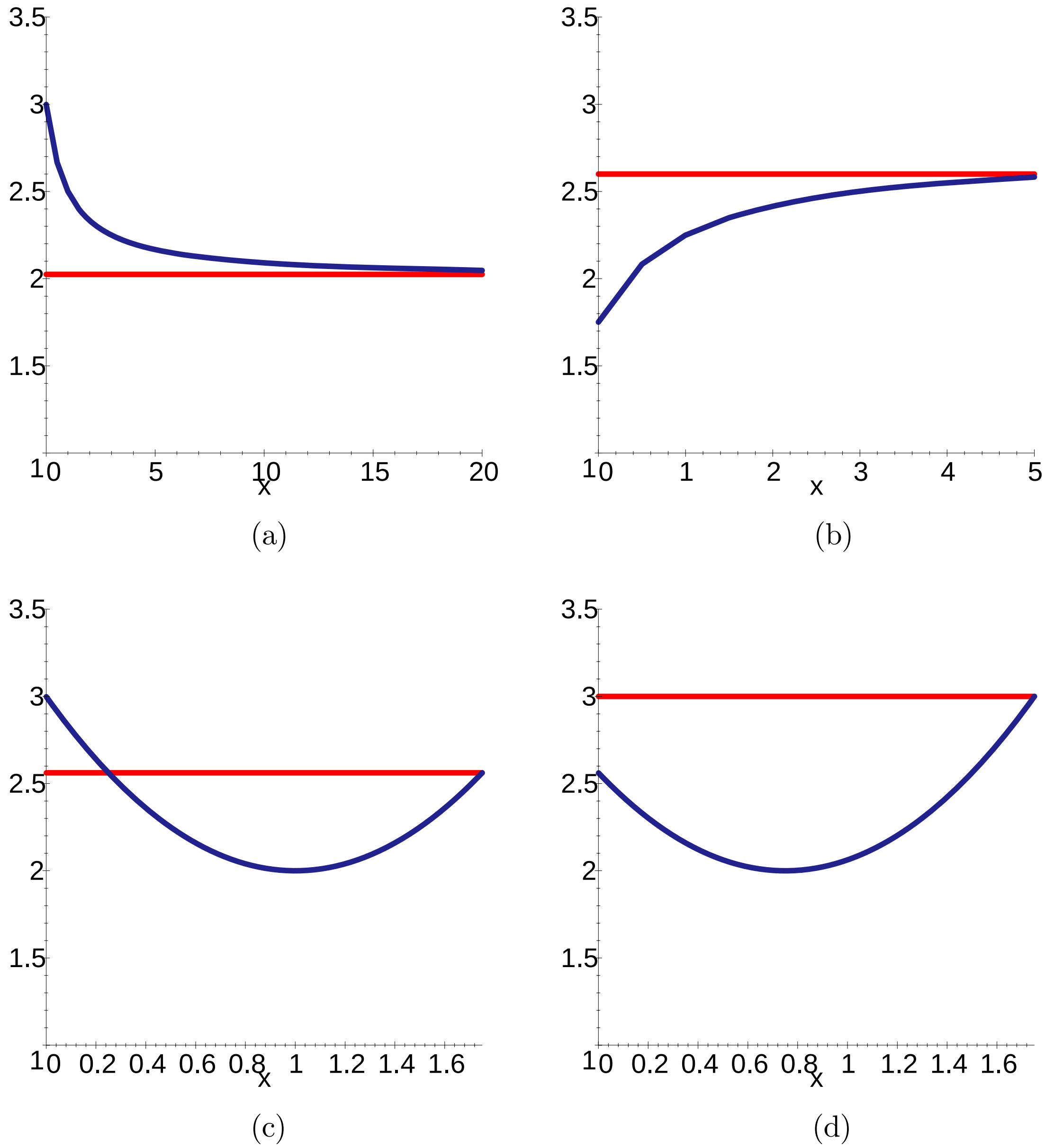
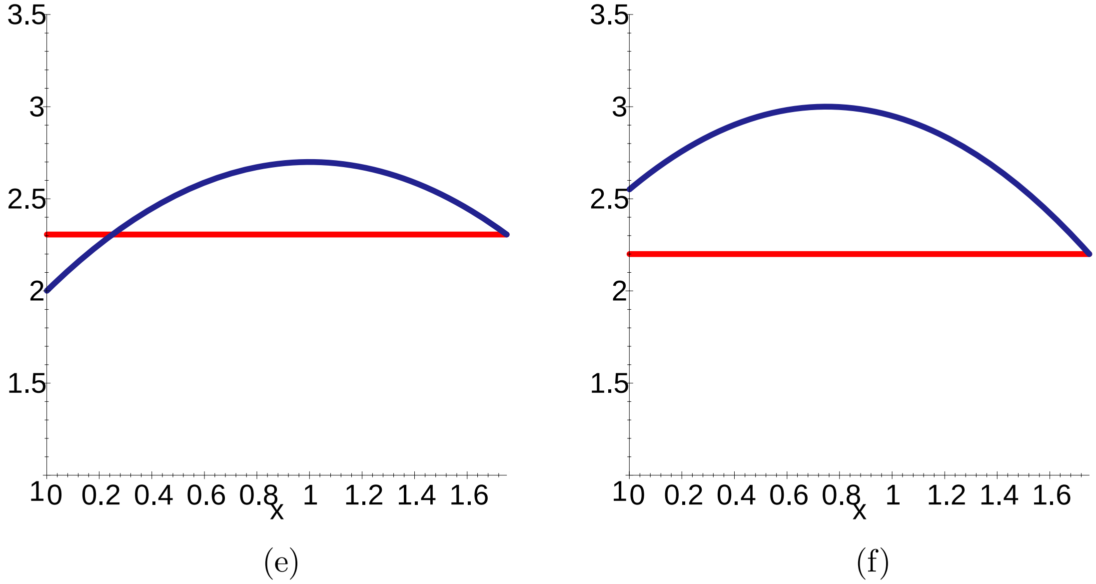


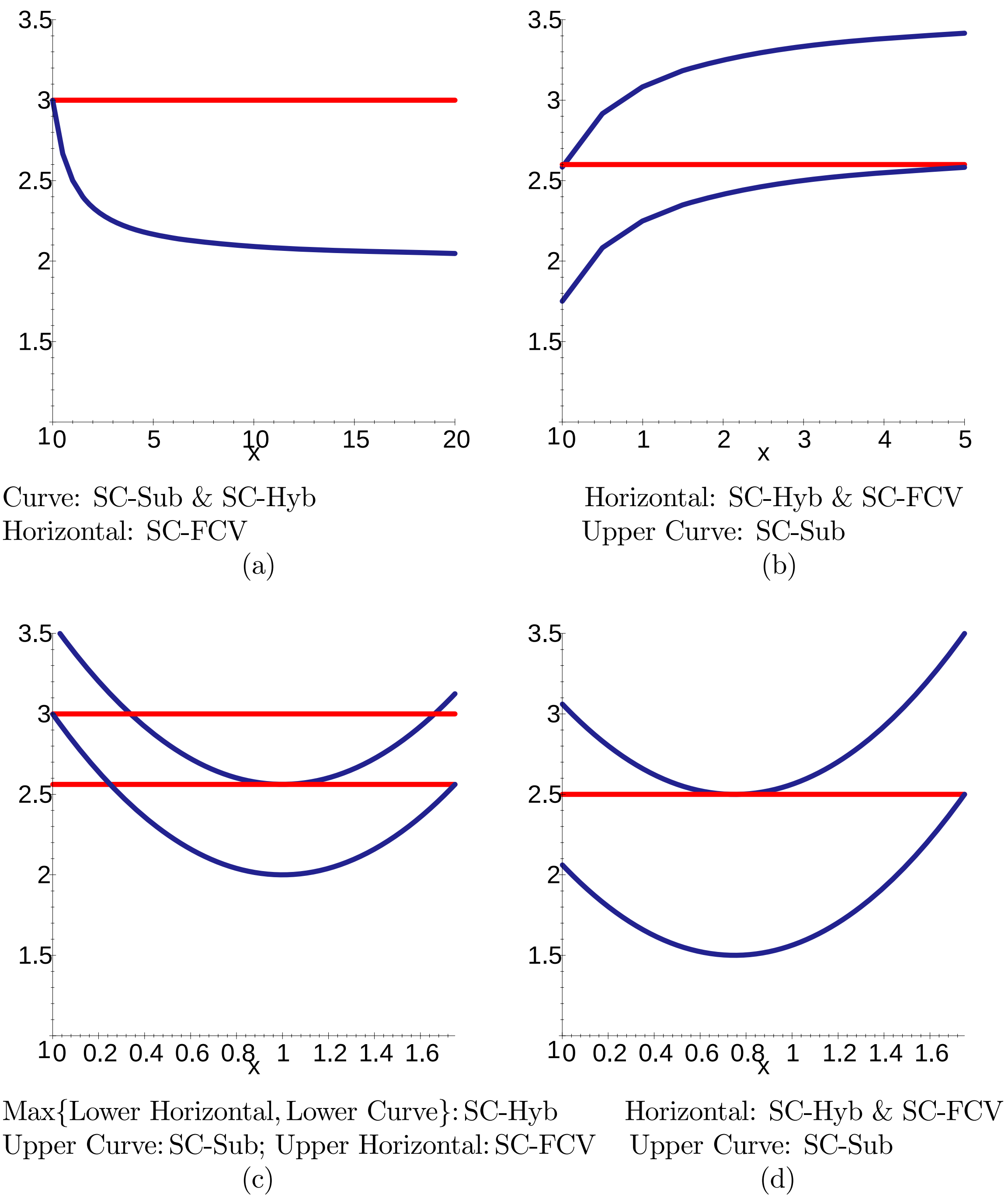
FIGURE B-1.—Hybrid, FCV, and Subsample Critical Values as a Function of $g \in H$: Hybrid = max{curved line, horizontal line}, FCV = horizontal line, Subsample = curved line



(e)

(f)

FIGURE B-1. (cont.).



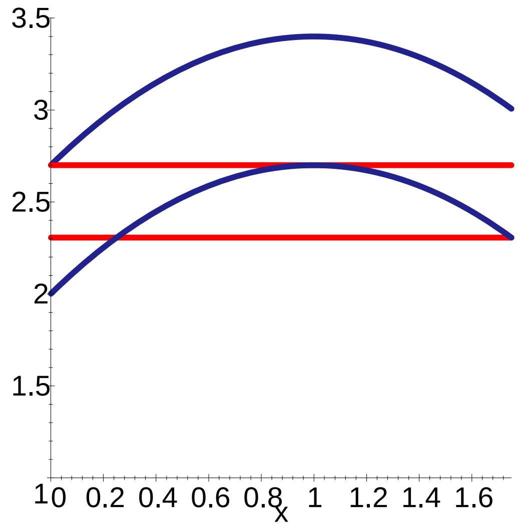
Curve: SC-Sub & SC-Hyb
Horizontal: SC-FCV
(a)

Horizontal: SC-Hyb & SC-FCV
Upper Curve: SC-Sub
(b)

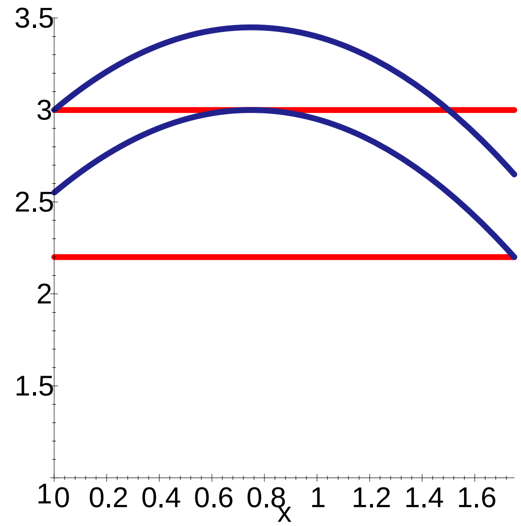
Max{Lower Horizontal, Lower Curve}: SC-Hyb
Upper Curve: SC-Sub; Upper Horizontal: SC-FCV
(c)

Horizontal: SC-Hyb & SC-FCV
Upper Curve: SC-Sub
(d)

FIGURE B-2.—Critical Values as a Function of $g \in H$ for SC-Sub, SC-FCV, and SC-Hyb Tests: In Each Panel the Lower Curve Is $c_g(1-\alpha)$ & the Lower Horizontal Is the FCV Critical Value



Upper Horizontal: SC-Hyb & SC-FCV
Upper Curve: SC-Sub
(e)



Upper Horizontal: SC-Hyb & SC-FCV
Upper Curve: SC-Sub
(f)

FIGURE B-2. (cont.).