ECONOM: 4875 Model 3G pp. 1–36 (col. fig: NIL)

ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Generic results for establishing the asymptotic size of confidence sets and tests ☆

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

[a] Cowles Foundation, Yale University, United States of America
[b] Department of Economics, University of Pennsylvania, United States of America
[c] Department of Economics, Pennsylvania State University, United States of America

A R T I C L E I N F O

Article history:
Available online xxxx

JEL classification:
C10
C12

Keywords:
Asymptotic size
Autoregressive model
Confidence set
Grid bootstrap
Identification
Inference
Test
Uniform
Weak instruments
Weak identification

A B S T R A C T

This paper provides a set of results that can be used to establish the asymptotic size and/or similarity in a uniform sense of confidence sets and tests. The results are generic in that they can be applied to a broad range of problems. They are most useful in scenarios where the pointwise asymptotic distribution of a test statistic is a discontinuous function of a parameter.

The results are illustrated in several examples. These are: (i) the conditional likelihood ratio test of Moreira (2003) for linear instrumental variables models with instruments that may be weak, extended to the case of heteroskedastic errors; (ii) the grid bootstrap confidence interval of Hansen (1999) for the sum of the AR coefficients in a kth order autoregressive model with unknown innovation distribution, and (iii) the standard quasi-likelihood ratio test in a nonlinear regression model where identification is lost when the coefficient on the nonlinear regressor is zero. In addition, as a simple running example, we consider a two-sided equal-tailed CI for the AR coefficient in an AR(1) model, which is a simplified version of the CI in Andrews and Guggenberger (2014).

1. Introduction

Throughout his long and productive career, Jean-Marie Dufour has worked on issues of uniformity and correct size of inferential procedures, in both finite-sample and asymptotic senses. This paper follows in this general line of research.

The objective of this paper is to provide results that can be used to convert asymptotic results under drifting sequences or subsequences of parameters into results that hold uniformly over a given parameter space. Such results can be used to establish the asymptotic size and asymptotic similarity of confidence sets (CS's) and tests. By definition, the asymptotic size of a CS or test is the limit of its finite-sample size. Also, by definition, the finite-sample size is a uniform concept, because it is the minimum coverage probability over a set of parameters/distributions for a CS and it is the maximum of the null rejection probability over a set for a test.

The size properties of CS's and tests are their most fundamental property. The asymptotic size is used to approximate the finite-sample size and typically it gives good approximations. On the other hand, it has been demonstrated repeatedly in the literature that pointwise asymptotics often provide very poor approximations to the finite-sample size in situations

☆ Andrews and Guggenberger gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-0751517, SES-1058376 and SES-1355504, and SES-1326827, respectively. The authors thank the co-editors, three referees, and Ming Li for helpful comments.
* Corresponding author.
E-mail address: donald.andrews@yale.edu (D.W.K. Andrews).

https://doi.org/10.1016/j.jeconom.2020.04.027
0304-4076/© 2020 Elsevier B.V. All rights reserved.

where the statistic of interest has a discontinuous pointwise asymptotic distribution. References are given below. Hence, it is useful to have available tools for establishing the asymptotic size of CS's and tests that are simple and easy to employ.

The results of this paper are useful in a wide variety of cases that have received attention recently in the econometrics literature. These are cases where the statistic of interest has a discontinuous pointwise asymptotic distribution. This means that the statistic has a different asymptotic distribution under different sequences of parameters/distributions that converge to the same parameter/distribution. Examples include: (i) time series models with unit roots, (ii) models in which identification fails to hold at some points in the parameter space, including weak instrumental variable (IV) scenarios, (iii) inference with moment inequalities, (iv) inference when a parameter may be at or near a boundary of the parameter space, and (v) post-model selection inference.

For example, in a simple autoregressive (AR) time series model $Y_i = \alpha + \rho Y_{i-1} + U_i$ for $i = 1, \ldots, n$, an asymptotic discontinuity arises for standard test statistics at the point $\rho = 1$. Standard statistics such as the least squares (LS) estimator and the LS-based t statistic have different asymptotic distributions as $n \to \infty$ if one considers a fixed sequence of parameters with $\rho = 1$ compared to a sequence of AR parameters $\rho_n = 1 - c/n^{\delta}$ for some constants $c \in R$ and $0 < \delta \leq 1$. But, in both cases, the limit of the AR parameter is one. Similarly, standard t tests in a linear IV regression model have asymptotic discontinuities at the reduced-form parameter value at which identification is lost.

The results of this paper show that to determine the asymptotic size and/or similarity of a CS or test it is sufficient to determine their asymptotic coverage or rejection probabilities under certain drifting subsequences of parameters/distributions. We start by providing general conditions for such results. Then, we give several sets of sufficient conditions for the general conditions that are easier to apply in practice.

No papers in the literature, other than the present paper, provide generic results of this sort. However, several papers in the literature provide uniform results for certain procedures in particular models or in some class of models. For example, Mikusheva (2007) uses a method based on almost sure representations to establish uniform properties of three types of confidence intervals (CI's) in autoregressive models with a root that may be near, or equal to, unity. Romano and Shaikh (2008, 2010) and Andrews and Guggenberger (2009c) provide uniform results for some subsampling CS's in the context of moment inequalities. Andrews and Soares (2010) provide uniform results for generalized moment selection CS's in the context of moment inequalities. Andrews and Guggenberger (2009a, 2010a) provide uniform results for subsampling, hybrid (combined fixed/subsampling critical values), and fixed critical values for a variety of cases. Romano and Shaikh (2012) provide uniform results for subsampling and the bootstrap that apply in some contexts.

The results in these papers are quite useful, but they have some drawbacks and are not easily transferable to different procedures in different models. For example, Mikusheva's (2007) approach using almost sure representations involves using an almost sure representation of the partial sum of the innovations and exploiting the linear form of the model to build up an approximating AR model based on Gaussian innovations. This approach cannot be applied (at least straightforwardly) to more complicated models with nonlinearities. Even in the linear AR model this approach does not seem to be conducive to obtaining results that are uniform over both the AR parameter ρ and the innovation distribution.

The approach of Romano and Shaikh (2008, 2010, 2012) applies to subsampling and bootstrap methods, but not to other methods of constructing critical values. The general theoretical results in Romano and Shaikh (2012) are based on a uniform comparison between the original and the subsampling/bootstrap distribution functions over the entire class of distributions. The distribution function comparison yields sufficient conditions for correct asymptotic size for any significance level. They are stronger than conditions in this paper which fixes the significance level. To establish uniform validity, Romano and Shaikh (2012) typically uses a contradiction argument, showing that one fails to find a subsequence of true models such that these distribution functions do not satisfy the sufficient condition. The approach cannot deliver the asymptotic size when it is different from the nominal size, as it often is in irregular models when subsampling or the standard bootstrap is applied. On the other hand, this paper breaks down the uniform result to some sequential limits and develops these limiting distribution functions. As such, this paper provides the asymptotic size, whether it is equal to the nominal size or not.

The approach of Andrews and Guggenberger (2009a,c, 2010a) applies to subsampling, hybrid, and fixed critical values, but is not designed for other methods. In particular, the general asymptotic results for CI's in Andrews and Guggenberger (2009c) apply to fixed and subsampling critical values. The results in Andrews and Guggenberger (2009c) do not cover bootstrap or conditional critical values. In this paper, we take the approach in Andrews and Guggenberger (2009a,c, 2010a) and generalize it so that it applies to a wide variety of cases, including any test statistic and any type of critical value(s). This approach is found to be quite flexible and easy to apply. It establishes the asymptotic size whether or not the asymptotic size is correct; it yields an explicit formula for asymptotic size; and it establishes asymptotic similarity when the latter holds.

We illustrate the results of the paper using several examples. The first example is a CI in a simple AR(1) model with AR parameter that may be close to, or equal to, one, and independent and identically-distributed innovations. This example is used as a running example to illustrate how the abstract general results are applied in practice. The CI considered is a two-sided equal-tailed CI that has correct asymptotic size and is asymptotically similar in a uniform sense. It is a simplification of a CI in Andrews and Guggenberger (2014).

The second example is a heteroskedasticity-robust version of the conditional likelihood ratio (CLR) test of Moreira (2003) for the linear IV regression model with one right-hand side endogenous variable. This test is designed to be robust to weak identification and heteroskedasticity. We show that the test has correct asymptotic size and is asymptotically similar in a uniform sense with errors that may be heteroskedastic and non-normal. Closely related tests are considered in Andrews et al. (2004, Section 9), Kleibergen (2005), Mikusheva (2010), Guggenberger (2012), Guggenberger et al. (2012b) and Andrews and Guggenberger (2019).

The third example is Hansen's (1999) grid bootstrap CI for the sum of the autoregressive coefficients in an AR(k) model. We show that the grid bootstrap CI has correct asymptotic size and is asymptotically similar in a uniform sense. We consider this example for comparative purposes because Mikusheva (2007) has established similar results. We show that our approach is relatively simple to employ—no almost sure representations are required. In addition, we obtain uniformity over different innovation distributions with little additional work.

The fourth example is a CI in a nonlinear regression model. In this model one loses identification in part of the parameter space because the nonlinearity parameter is unidentified when the coefficient on the nonlinear regressor is zero. We consider standard quasi-likelihood ratio (QLR) CI's. We show that such CI's do not necessarily have correct asymptotic size and are not asymptotically similar typically. We provide expressions for the degree of asymptotic size distortion and the magnitude of asymptotic non-similarity. These results make use of some results in Andrews and Cheng (2012b).

The method of this paper also has been used to establish the asymptotic size of tests and/or CS's in Andrews and Cheng (2012a, 2013, 2014) for t, Wald, and QLR statistics in models that exhibit lack of identification at some points in the parameter space; Andrews and Guggenberger (2014) for the AR(1) parameter in an AR(1) model with conditional heteroskedasticity; Andrews and Guggenberger (2017, 2019) for various identification-robust tests in nonlinear moment condition models; Cheng (2015) for nonlinear regression models with multiple sources of weak identification; Cheng et al. (2019) for the risk of an averaging GMM estimator that incorporates valid and invalid moment conditions; and Feir et al. (2016) for fuzzy regression discontinuity models with weak identification.

We note that some of the results of this paper do not hold in scenarios in which the parameter that determines whether one is at a point of discontinuity is infinite dimensional. This arises in tests of stochastic dominance and CS's based on conditional moment inequalities, e.g., see Andrews and Shi (2013, 2014) and Linton et al. (2010).

Selected references in the literature regarding uniformity issues in the models discussed above include the following: for unit roots, Bobkowski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987), Stock (1991), Park (2002), Giraitis and Phillips (2006), Phillips and Magdalinos (2007), and Andrews and Guggenberger (2012); for weak identification due to weak IV's, Staiger and Stock (1997), Stock and Wright (2000), Moreira (2003), Kleibergen (2005), Guggenberger et al. (2012a), Guggenberger et al. (2012b, 2019); for weak identification in other models, Andrews and Cheng (2012a, 2013, 2014), Qu (2014), Andrews and Mikusheva (2015, 2016), Cox (2016), and Han and McCloskey (2019); for parameters near a boundary, Chernoff (1954), Self and Liang (1987), Shapiro (1989), Geyer (1994), Andrews (1999, 2001, 2002), Andrews and Guggenberger (2010b), and McCloskey (2017); and for post-model selection inference, Kabaila (1995), Leeb and Pötscher (2005), Leeb (2006), Andrews and Guggenberger (2009a,b) and McCloskey (2017).

This paper is organized as follows. Section 2 provides the generic asymptotic size and similarity results. Section 3 gives the uniformity results for the CLR test in the linear IV regression model. Section 4 provides the results for Hansen's (1999) grid bootstrap in the AR(k) model. Section 5 gives the uniformity results for the quasi-LR test in the nonlinear regression model. An Appendix provides proofs of some results used in the examples given in Sections 3–5.

2. Determination of asymptotic size and similarity

2.1. General results

This subsection provides the most general results of the paper. We state a theorem that is useful in a wide variety of circumstances when calculating the asymptotic size of a sequence of CS's or tests. It relies on the properties of the CS's or tests under drifting sequences or subsequences of true distributions.

Let $\{CS_n : n \geq 1\}$ be a sequence of CS's for a parameter $r(\lambda)$, where λ indexes the true distribution of the observations, the parameter space for λ is some space Λ, and $r(\lambda)$ takes values in some space $\mathcal{R}$. In many cases, λ is of the form $\lambda = (\lambda_1, \ldots, \lambda_q, \lambda_{q+1})'$, where $\lambda_j \in R\, \forall j \leq q$ are parameters and λ_{q+1} is infinite-dimensional, such as the distribution of an error term. Let $CP_n(\lambda)$ denote the coverage probability of CS_n under λ.

An example of a nominal $1-\alpha$ CS for $r(\lambda)$ is

$$CS_n = \{v : T_n(v) \leq c_{n,1-\alpha}(v)\}, \tag{2.1}$$

where $T_n(v)$ is a test statistic, such as a t, Wald, or QLR statistic, and $c_{n,1-\alpha}(v)$ is a critical value for testing $H_0 : r(\lambda) = v$. Critical values considered in this paper may depend on the null value v of $r(\lambda)$, as well as on the sample size n and the data. The coverage probability of this CS for $r(\lambda)$ is

$$CP_n(\lambda) \equiv P_\lambda(r(\lambda) \in CS_n) = P_\lambda(T_n(r(\lambda)) \leq c_{n,1-\alpha}(r(\lambda))). \tag{2.2}$$

The *exact size* and *asymptotic size* of CS_n are denoted by

$$ExSz_n \equiv \inf_{\lambda \in \Lambda} CP_n(\lambda) \text{ and } AsySz \equiv \liminf_{n\to\infty}\ ExSz_n, \tag{2.3}$$

respectively. A nominal $1-\alpha$ CS has correct asymptotic level $1-\alpha$ if

$$AsySz \geq 1-\alpha. \tag{2.4}$$

By definition, a CS is *similar* in finite samples if $CP_n(\lambda)$ does not depend on λ for $\lambda \in \Lambda$. In other words, a CS is similar if

$$\inf_{\lambda\in\Lambda} CP_n(\lambda) = \sup_{\lambda\in\Lambda} CP_n(\lambda). \tag{2.5}$$

We say that a sequence of CS's $\{CS_n : n \geq 1\}$ is *asymptotically similar* (in a uniform sense) if

$$\liminf_{n\to\infty} \inf_{\lambda\in\Lambda} CP_n(\lambda) = \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} CP_n(\lambda). \tag{2.6}$$

Define the asymptotic maximum coverage probability of $\{CS_n : n \geq 1\}$ by

$$AsyMaxCP \equiv \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} CP_n(\lambda). \tag{2.7}$$

Then, a sequence of CS's is asymptotically similar if $AsySz = AsyMaxCP$.

We now introduce an example that is designed to illustrate how the general results given below are applied in practice.

Example 1. We consider an equal-tailed two-sided confidence interval (CI) for the autoregressive parameter (AR) ρ in an AR(1) model in which ρ may be close to, or equal to, one. The CI is shown to have correct asymptotic size and be asymptotically similar. We consider a model with i.i.d. innovations. Andrews and Guggenberger (2014) provide a version of the CI that applies with conditionally heteroskedastic innovations. We use the unobserved components representations of the AR(1) model. The observed time series $\{Y_i : i = 0, \ldots, n\}$ is based on a latent no-intercept AR(1) time series $\{Y_i^* : i = 0, \ldots, n\}$:

$$\begin{aligned} Y_i &= \alpha + Y_i^*, \\ Y_i^* &= \rho Y_{i-1}^* + U_i, \text{ for } i = 1, \ldots, n, \end{aligned} \tag{2.8}$$

where $\rho \in [-1+\varepsilon, 1]$ for some $0 < \varepsilon < 2$, $\{U_i : i = \cdots, 0, 1, \ldots\}$ are i.i.d. with distribution F and mean 0. The distribution of Y_0^* is the distribution that yields strict stationarity for $\{Y_i^* : i \leq n\}$ when $\rho < 1$. That is, $Y_0^* = \sum_{j=0}^{\infty} \rho^j U_{-j}$ when $\rho < 1$. When $\rho = 1$, Y_0^* is arbitrary. Note that the model can be rewritten as

$$Y_i = \widetilde{\alpha} + \rho Y_{i-1} + U_i, \text{ where } \widetilde{\alpha} \equiv \alpha(1-\rho) \tag{2.9}$$

for $i = 1, \ldots, n$.[1]

The CI is obtained by inverting a test of the null hypothesis that the true value is ρ. The test uses the t statistic

$$T_n(\rho) \equiv \frac{n^{1/2}(\widehat{\rho}_n - \rho)}{\widehat{\sigma}_n}, \tag{2.10}$$

where $\widehat{\rho}_n$ is the LS estimator from the regression of Y_i on a constant and Y_{i-1} and $\widehat{\sigma}_n^2$ is a variance estimator defined immediately below. Let Y, U, X_1, and X_2 be n-vectors with ith elements given by Y_i, U_i, Y_{i-1}, and 1, respectively. Define $M_{X_2} = I_n - X_2(X_2'X_2)^{-1}X_2'$ (and analogously for vectors other than X_2) and $\widehat{U} = M_{M_{X_2}X_1} M_{X_2} Y$. The OLS estimator and variance estimator are defined as

$$\widehat{\rho}_n = \left(X_1'M_{X_2}X_1\right)^{-1} X_1'M_{X_2}Y \text{ and } \widehat{\sigma}_n^2 = \left(n^{-1}X_1'M_{X_2}X_1\right)^{-1} n^{-1}\widehat{U}'\widehat{U}. \tag{2.11}$$

The parameter space for (ρ, F) is

$$\begin{aligned} \Lambda \equiv \{\lambda = (\rho, F) : \rho \in [-1+\varepsilon, 1], \quad &\{U_i\colon i = 0, \pm 1, \pm 2, \ldots\} \text{ are i.i.d.} \\ &\text{with distribution } F, \quad E_F U_i = 0, \quad E_F|U_i|^4 \leq M, \text{ and } E_F U_i^2 \geq \delta\}, \end{aligned} \tag{2.12}$$

for some constants $0 < \varepsilon < 2$, $M < \infty$, and $\delta > 0$.

The critical values used in the construction of the CI are based on the asymptotic distributions of the test statistic under drifting sequences $\{\lambda_n = (\rho_n, F_n) \in \Lambda : n \geq 1\}$ of AR parameters ρ_n and distributions F_n that satisfy

$$n(1-\rho_n) \to h \in H = [0, \infty]. \tag{2.13}$$

[1] The advantage of writing the model as in (2.8) becomes clear here. The case $\rho = 1$ and $\widetilde{\alpha} \neq 0$ is automatically ruled out by model (2.8). This is a case where Y_i is dominated by a deterministic trend and the LS estimator of ρ converges at rate $n^{3/2}$.

To describe the asymptotic distribution, let $W(\cdot)$ be a standard Brownian motion on [0, 1]. Let Z be a standard normal random variable that is independent of $W(\cdot)$. Define

$$I_h(r) \equiv \int_0^r \exp(-(r-s)h)dW(s), \quad I_h^*(r) \equiv I_h(r) + \frac{1}{\sqrt{2h}}\exp(-hr)Z \text{ for } h > 0,$$
$$I_h^*(r) \equiv W(r) \text{ for } h = 0, \text{ and } I_{D,h}^*(r) \equiv I_h^*(r) - \int_0^1 I_h^*(s)ds. \tag{2.14}$$

Lemma A.1 in the Appendix shows that, under any such sequence $\{\lambda_n\}$, we have

$$T_n(\rho_n) \to_d J_h \text{ for } h \in H, \tag{2.15}$$

where J_h is defined as follows. For $h = \infty$, J_h is the $N(0, 1)$ distribution, and for $h \in [0, \infty)$, J_h is the distribution of

$$\int_0^1 I_{D,h}^*(r)dW(r) \Big/ \left(\int_0^1 I_{D,h}^*(r)^2 dr\right)^{1/2}. \tag{2.16}$$

For $\alpha \in (0, 1)$, let $c_h(1-\alpha)$ denote the $(1-\alpha)$-quantile of J_h. The second component of $I_h^*(r)$ in (2.14) is due to the stationary start-up of the AR(1) process when $\rho < 1$, as in Elliott (1999), Elliott and Stock (2001), Müller and Elliott (2003), and Andrews and Guggenberger (2009a, 2012). Giraitis and Phillips (2006) and Phillips and Magdalinos (2007) provide similar results for the LS estimator for the case $h = \infty$.

The nominal $1 - \alpha$ equal-tailed two-sided CI for ρ is

$$CI_n = \{\rho \in [-1+\varepsilon, 1] : c_{h_n}(\alpha/2) \le T_n(\rho) \le c_{h_n}(1-\alpha/2) \text{ for } h_n = n(1-\rho)\}. \tag{2.17}$$

The CI CI_n can be calculated by taking a fine grid of values $\rho \in [-1+\varepsilon, 1]$ and comparing $T_n(\rho)$ to $c_{h_n}(\alpha/2)$ and $c_{h_n}(1-\alpha/2)$, where $h_n = n(1-\rho)$. Tables of values of $c_h(\alpha/2)$ and $c_h(1-\alpha/2)$ are given in Andrews and Guggenberger (2014). Given these values, calculation of CI_n is simple and fast. □

Now we return to the general results. For a sequence of constants $\{\kappa_n : n \ge 1\}$, let $\kappa_n \to [\kappa_{1,\infty}, \kappa_{2,\infty}]$ denote that $\kappa_{1,\infty} \le \liminf_{n\to\infty} \kappa_n \le \limsup_{n\to\infty} \kappa_n \le \kappa_{2,\infty}$. All limits are as $n \to \infty$ unless stated otherwise.

We use the following assumptions:

Assumption A1. For any sequence $\{\lambda_n \in \Lambda : n \ge 1\}$ and any subsequence $\{w_n\}$ of $\{n\}$ there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that

$$CP_{p_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)] \tag{2.18}$$

for some $CP^-(h), CP^+(h) \in [0, 1]$ and some h in an index set H.[2]

Assumption A1 specifies the limiting behavior of coverage probabilities along certain (sub)sequences. One could associate each (λ_n, w_n, p_n) (sub)sequence with a value h, but typically a value of h is associated with some collection of (λ_n, w_n, p_n) (sub)sequences that have similar limiting behavior. For example, in Example 1, if for some subsequence $\{p_n\}$ of $\{w_n\}$, $p_n(1-\rho_{p_n}) \to h$, then (λ_n, w_n, p_n) is indexed by h and (2.15) and (2.16) describe the common limiting behavior (along $\{p_n\}$) for all (λ_n, w_n, p_n) associated with h. Eq. (2.24) and Assumption B below further illuminate the connection between (sub)sequences and indices h.

Assumption A1 defines the index set H, as well as $CP^-(h)$ and $CP^+(h)$ for $h \in H$. For some choices of $CP^-(h)$ and $CP^+(h)$, such as 0 and 1, Assumption A1 can always be made to hold. But, these lead to uninformative bounds on asymptotic size. Assumption A1 leads to more informative bounds the smaller is $CP^+(h) - CP^-(h)$. Hence, one aims to verify Assumption A1 with $CP^+(h) - CP^-(h)$ equal to zero, or as close to zero as possible, for each $h \in H$.

Example 1 (*Cont.*).In this example, Assumption A1 holds with $\lambda = (\rho, F)$, Λ defined in (2.12), $h \in H = [0, \infty]$, and $CP^-(h) = CP^+(h) = 1 - \alpha$ by Lemma A.6 in the Appendix. In this example, $CP^-(h)$ and $CP^+(h)$ do not depend on h, but in other examples they do, e.g., see the nonlinear regression example in Section 5. □

---

[2] It is not the case that Assumption A1 can be replaced by the following simpler condition just by re-indexing the parameters: **Assumption A1†**. For any sequence $\{\lambda_n \in \Lambda : n \ge 1\}$, there exists a subsequence $\{p_n\}$ of $\{n\}$ for which (2.18) holds.

The flawed re-indexing argument goes as follows: Let $\{\lambda_{w_n} : n \ge 1\}$ be an arbitrary subsequence of $\{\lambda_n \in \Lambda : n \ge 1\}$. We want to use Assumption A1† to show that there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that $CP_{p_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)]$. Define a new sequence $\{\lambda_n^* \in \Lambda : n \ge 1\}$ by $\lambda_n^* = \lambda_{w_n}$. By Assumption A1†, there exists a subsequence $\{k_n\}$ of $\{n\}$ such that $CP_{k_n}(\lambda_{k_n}^*) \to [CP^-(h), CP^+(h)]$, which looks close to the desired result. However, in terms of the original subsequence $\{\lambda_{w_n} : n \ge 1\}$ of interest, this gives $CP_{k_n}(\lambda_{w_{k_n}}) \to [CP^-(h), CP^+(h)]$ because $\lambda_{k_n}^* = \lambda_{w_{k_n}}$. Defining $p_n = w_{k_n}$, we obtain a subsequence $\{p_n\}$ of $\{w_n\}$ for which $CP_{k_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)]$. This is not the desired result because the subscript k_n on $CP_{k_n}(\cdot)$ is not the desired subscript.

In the following assumptions, the set H is as in Assumption A1.

Assumption A2. $\forall h \in H$, there exists a subsequence $\{p_n\}$ of $\{n\}$ and a sequence $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ such that (2.18) holds.[3]

Assumption C1. $CP^-(h_L) = CP^+(h_L)$ for some $h_L \in H$ such that $CP^-(h_L) = \inf_{h\in H} CP^-(h)$.

Assumption C2. $CP^-(h_U) = CP^+(h_U)$ for some $h_U \in H$ such that $CP^+(h_U) = \sup_{h\in H} CP^+(h)$.

Assumption S. $CP^-(h) = CP^+(h) = CP \ \forall h \in H$, for some constant $CP \in [0, 1]$.

In applications, Assumptions C1 and C2 typically hold because $CP^-(h) = CP^+(h) \ \forall h \in H$.

In practice, Assumptions C1 and C2 typically are continuity conditions (hence, C stands for continuity). This is because given any subsequence $\{w_n\}$ of $\{n\}$ one usually can choose a subsequence $\{p_n\}$ of $\{w_n\}$ such that $CP_{p_n}(\lambda_{p_n})$ has a well-defined limit $CP(h)$, in which case $CP^-(h) = CP^+(h) = CP(h)$. However, this is not possible in some troublesome cases. For example, suppose CS_n is defined by inverting a test that is based on a test statistic and a fixed critical value and the asymptotic distribution function of the test statistic under $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ has a discontinuity at the critical value. Then, it is usually only possible to show $CP_{p_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)]$ for some $CP^-(h) < CP^+(h)$. Assumption A1 allows for troublesome cases such as this. Assumption C1 holds if for at least one value $h_L \in H$ for which $CP^-(h_L) = \inf_{h\in H} CP^-(h)$ such a troublesome case does not arise. Assumption C2 is analogous.

Assumption S (where S stands for similarity) requires that the asymptotic coverage probabilities of the CS's in Assumptions A1 and A2 do not depend on the particular sequence of parameter values considered. When this assumption holds, one can establish asymptotic similarity of the CS's. When it fails, the CS's are not asymptotically similar.

Example 1 (*cont.*)**.** Assumptions A2, C1, C2, and S all hold in this example (by the results stated above for this example). □

The most general result of the paper is the following.

Theorem 2.1. *The confidence sets $\{CS_n : n \geq 1\}$ satisfy the following results.*

(a) *Under Assumption* A1, $\inf_{h\in H} CP^-(h) \leq AsySz \leq AsyMaxCP \leq \sup_{h\in H} CP^+(h)$.

(b) *Under Assumptions* A1 *and* A2, $AsySz \in \left[\inf_{h\in H} CP^-(h), \inf_{h\in H} CP^+(h)\right]$ *and* $AsyMaxCP \in \left[\sup_{h\in H} CP^-(h), \sup_{h\in H} CP^+(h)\right]$.

(c) *Under Assumptions* A1, A2, *and* C1, $AsySz = \inf_{h\in H} CP^-(h) = \inf_{h\in H} CP^+(h)$.

(d) *Under Assumptions* A1, A2, *and* C2, $AsyMaxCP = \sup_{h\in H} CP^-(h) = \sup_{h\in H} CP^+(h)$.

(e) *Under Assumptions* A1 *and* S, $AsySz = AsyMaxCP = CP$.

Comments. 1. Theorem 2.1 provides bounds on, and explicit expressions for, *AsySz* and *AsyMaxCP*. Theorem 2.1(e) provides sufficient conditions for asymptotic similarity of CS's.

2. The parameter space Λ may depend on n without affecting the results. Allowing Λ to depend on n allows one to cover local violations of some assumptions, as in Guggenberger (2012).

3. The results of Theorem 2.1 hold when the parameter that determines whether one is at a point of discontinuity (of the pointwise asymptotic coverage probabilities) is finite or infinite dimensional or if no such point or points exist.

4. Theorem 2.1 (and other results below) apply to CS's, rather than tests. However, if the following changes are made, then the results apply to tests. One replaces (i) the sequence of CS's $\{CS_n : n \geq 1\}$ by a sequence of tests $\{\phi_n : n \geq 1\}$ of some null hypothesis, (ii) "*CP*" by "*RP*" (which abbreviates null rejection probability), (iii) *AsyMaxCP* by *AsyMinRP* (which abbreviates asymptotic minimum null rejection probability), and (iv) "inf " by "sup" throughout (including in the definition of exact size). In addition, (v) one takes Λ to be the parameter space under the null hypothesis rather than the entire parameter space.[4] The proofs go through with the same changes provided the directions of inequalities are reversed in various places.

5. The definitions of *similar on the boundary* (of the null hypothesis) of a test in finite samples and asymptotically are the same as those for similarity of a test, but with Λ denoting the boundary of the null hypothesis, rather than the entire null hypothesis. Theorem 2.1 can be used to establish the asymptotic similarity on the boundary (in a uniform sense) of a test by defining Λ in this way and making the changes described in Comment 4.

Proof of Theorem 2.1. First we establish part (a). To show $AsySz \geq \inf_{h\in H} CP^-(h)$, let $\{\lambda_n \in \Lambda : n \geq 1\}$ be a sequence such that $\liminf_{n\to\infty} CP_n(\lambda_n) = \liminf_{n\to\infty} \inf_{\lambda\in\Lambda} CP_n(\lambda) (= AsySz)$. Such a sequence always exists. Let $\{w_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} CP_{w_n}(\lambda_{w_n})$ exists and equals *AsySz*. Such a sequence always exists. By Assumption A1, there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that (2.18) holds for some $h \in H$. Hence,

$$AsySz = \lim_{n\to\infty} CP_{p_n}(\lambda_{p_n}) \geq CP^-(h) \geq \inf_{h\in H} CP^-(h). \tag{2.19}$$

[3] The following conditions are equivalent to Assumptions A1 and A2, respectively:
Assumption A1-alt. For any sequence $\{\lambda_n \in \Lambda : n \geq 1\}$ and any subsequence $\{w_n\}$ of $\{n\}$ such that $CP_{w_n}(\lambda_{w_n})$ converges, $\lim_{n\to\infty} CP_{w_n}(\lambda_{w_n}) \in [CP^-(h), CP^+(h)]$ for some $CP^-(h), CP^+(h) \in [0, 1]$ and some h in an index set H. **Assumption A2-alt.** $\forall h \in H$, there exists a subsequence $\{w_n\}$ of $\{n\}$ and a sequence $\{\lambda_{w_n} \in \Lambda : n \geq 1\}$ such that $CP_{w_n}(\lambda_{w_n})$ converges and $\lim_{n\to\infty} CP_{w_n}(\lambda_{w_n}) \in [CP^-(h), CP^+(h)]$.

[4] The null hypothesis and/or the parameter space Λ can be fixed or drifting with n.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

The proof that $AsyMaxCP \leq \sup_{h\in H} CP^+(h)$ is analogous to the proof just given with $AsySz$, $\inf_{h\in H}$, $\inf_{\lambda\in\Lambda}$, $CP^-(h)$, and $\liminf_{n\to\infty}$ replaced by $AsyMaxCP$, $\sup_{h\in H}$, $\sup_{\lambda\in\Lambda}$, $CP^+(h)$, and $\limsup_{n\to\infty}$, respectively. The inequality $AsySz \leq AsyMaxCP$ holds immediately given the definitions of these two quantities, which completes the proof of part (a).

Given part (a), to establish the $AsySz$ result of part (b) it suffices to show that

$$AsySz \leq CP^+(h) \;\; \forall h \in H. \tag{2.20}$$

Given any $h \in H$, let $\{p_n\}$ and $\{\lambda_{p_n}\}$ be as in Assumption A2. Then, we have:

$$AsySz = \liminf_{n\to\infty} \inf_{\lambda\in\Lambda} CP_n(\lambda) \leq \liminf_{n\to\infty} \inf_{\lambda\in\Lambda} CP_{p_n}(\lambda) \leq \liminf_{n\to\infty} CP_{p_n}(\lambda_{p_n}) \leq CP^+(h). \tag{2.21}$$

This proves the $AsySz$ result of part (b). The $AsyMaxCP$ result of part (b) is proved analogously with $\liminf_{n\to\infty} \inf_{\lambda\in\Lambda}$ replaced by $\limsup_{n\to\infty} \sup_{\lambda\in\Lambda}$.

Part (c) of the Theorem follows from part (b) plus $\inf_{h\in H} CP^-(h) = \inf_{h\in H} CP^+(h)$. The latter holds because

$$\inf_{h\in H} CP^-(h) = CP^-(h_L) = CP^+(h_L) \geq \inf_{h\in H} CP^+(h) \tag{2.22}$$

by Assumption C1, and $\inf_{h\in H} CP^-(h) \leq \inf_{h\in H} CP^+(h)$ because $CP^-(h) \leq CP^+ \; \forall h \in H$ by Assumption A2 (which guarantees the existence of $CP^-(h)$ and $CP^+(h) \; \forall h \in H$). Part (d) of the Theorem holds by an analogous argument as for part (c) using Assumption C2 in place of Assumption C1.

Part (e) follows from part (a) because Assumption S implies that $\inf_{h\in H} CP^-(h) = \sup_{h\in H} CP^+(h) = CP$. □

2.2. Sufficient conditions

In this subsection, we provide several sets of sufficient conditions for Assumptions A1 and A2. They show how Assumptions A1 and A2 can be verified in practice. These sufficient conditions apply when the parameter that determines whether one is at a point of discontinuity (of the pointwise asymptotic coverage probabilities) is finite dimensional, but not infinite dimensional.

First we introduce a condition, Assumption B, that is sufficient for Assumptions A1 and A2. Let $\{h_n(\lambda) : n \geq 1\}$ be a sequence of functions on Λ, where

$$h_n(\lambda) = (h_{n,1}(\lambda), \ldots, h_{n,J}(\lambda), h_{n,J+1}(\lambda))', \tag{2.23}$$

$h_{n,j}(\lambda) \in R \; \forall j \leq J$, and $h_{n,J+1}(\lambda) \in \mathcal{T}$ for some compact metric space $\mathcal{T}$.[5] In many cases, $h_{n,1}(\lambda)$ is of the form $d_{n,1}(a + b\lambda_1)$ for some constants $\{d_{n,1}\}$, such as $\{n^{1/2}\}$ or $\{n\}$, some constants a and b, and $h_{n,j}(\lambda) = \lambda_j$ for $j = 2, \ldots, J$. For example, in the AR(1) Example, $d_{n,1} = n$, $a = 1$, $b = -1$, and $a + b\lambda_1 = 1 - \rho$. In many other examples, $d_{n,1} = n^{1/2}$, $a = 0$, and $b = 1$. If an infinite-dimensional parameter does not arise in the model of interest, or if such a parameter arises but does not affect the asymptotic coverage probabilities of the CS's under the drifting sequences $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ considered here, then the last element $h_{n,J+1}(\lambda)$ of $h_n(\lambda)$ is not needed and can be omitted from the definition of $h_n(\lambda)$. For example, this is the case in all of the examples considered in Andrews and Guggenberger (2009a, 2010a, 2009b).

Suppose the CS's $\{CS_n : n \geq 1\}$ depend on a test statistic and a possibly data-dependent critical value. Then, the function $h_n(\lambda)$ is chosen so that if $h_n(\lambda_n)$ converges to some limit, say h (whose elements might include $\pm\infty$), for some sequence of parameters $\{\lambda_n\}$, then the test statistic and critical value converge in distribution to some limit distributions that may depend on h. In short, $h_n(\lambda)$ is chosen so that convergence of $h_n(\lambda_n)$ yields convergence of the test statistic and critical value. See the examples below for illustrations.

Example 1 (*cont.*)**.** In this example, one takes $h_n(\lambda) = n(1 - \rho)$, $J = 1$, $H = [0, \infty]$, and the component $h_{n,J+1}(\lambda)$ does not arise. □

For other examples, Andrews and Cheng (2012a, 2013, 2014) analyze CS's and tests constructed using t, Wald, and quasi-LR test statistics and (possibly data-dependent) critical values based on a criterion function that depends on parameters $(\beta', \zeta', \pi')'$, where the parameter $\pi \in R^{d_\pi}$ is not identified when $\beta = 0 \in R^{d_\beta}$, and the parameters $(\beta', \zeta')' \in R^{d_\beta + d_\zeta}$ are always identified. The distribution that generates the data is indexed by $\gamma = (\beta', \zeta', \pi', \phi)'$, where the parameter $\phi \in \mathcal{T}$ and $\mathcal{T}$ is some compact metric space. In this scenario, one takes $\lambda = (\|\beta\|, \beta'/\|\beta\|, \zeta', \pi', \phi)'$ and $h_n(\lambda) = (n^{1/2}\|\beta\|, \|\beta\|, \beta'/\|\beta\|, \zeta', \pi', \phi)'$, where if $\beta = 0$, then $\beta/\|\beta\| = 1_{d_\beta}/\|1_{d_\beta}\|$ for $1_{d_\beta} = (1, \ldots, 1)' \in R^{d_\beta}$. (If ϕ does not affect the limit distributions of the test statistic and critical value, then it can be dropped from $h_n(\lambda)$ and $\mathcal{T}$ does not need to be compact.)

In Andrews and Guggenberger (2009a, 2010a,b), which considers subsampling and m out of n bootstrap CI's and tests with subsample or bootstrap size m_n and uses a parameter $\gamma = (\gamma_1', \gamma_2', \gamma_3')'$, one takes $\lambda = \gamma$ and $h_n(\lambda) = (n^r\gamma_1, m_n^r\gamma_1, \gamma_2')'$

[5] For notational simplicity, we stack J real-valued quantities and one $\mathcal{T}$-valued quantity into the vector $h_n(\lambda)$. Theorem 2.2 and Lemma 2.1 hold if $\mathcal{T}$ is a compact topological space, but Lemma 2.2 uses the stronger assumption that $\mathcal{T}$ is a compact metric space. We are not aware of any examples where the metric space assumption is restrictive, although the compactness assumption can be.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

(in the case of a test), where $r = 1/2$ in most applications (other than unit root models), $\gamma_1 \in R^p$, $\gamma_2 \in R^q$, $\gamma_3 \in \mathcal{T}$, and $\mathcal{T}$ is some arbitrary (possibly infinite-dimensional) space.

Define

$$H \equiv \{h \in (R \cup \{\pm\infty\})^J \times \mathcal{T} : h_{p_n}(\lambda_{p_n}) \to h \text{ for some subsequence } \{p_n\} \text{ of } \{n\} \text{ and some sequence } \{\lambda_{p_n} \in \Lambda : n \geq 1\}\}. \quad (2.24)$$

Assumption B. For any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ for which $h_{p_n}(\lambda_{p_n}) \to h \in H$, $CP_{p_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)]$ for some $CP^-(h), CP^+(h) \in [0, 1]$.

Theorem 2.2. *Assumption* B *implies Assumptions* A1 *and* A2.

The proof of Theorem 2.2 is given in Section 2.3.

The parameter λ and function $h_n(\lambda)$ in Assumption B typically are of the following form: (i) For all $\lambda \in \Lambda$, $\lambda = (\lambda_1, \ldots, \lambda_q, \lambda_{q+1})'$, where $\lambda_j \in R \, \forall j \leq q$ and λ_{q+1} belongs to some infinite-dimensional metric space. Often λ_{q+1} is the distribution of some random variable or vector, such as the distribution of one or more error terms or the distribution of the observable variables or some function of them.

(ii) $h_n(\lambda) = (h_{n,1}(\lambda), \ldots, h_{n,J}(\lambda), h_{n,J+1}(\lambda))'$ for $\lambda \in \Lambda$ is of the form:

$$h_{n,j}(\lambda) = \begin{cases} d_{n,j}\lambda_j & \text{for } j = 1, \ldots, J_* \\ m_j(\lambda) & \text{for } j = J_* + 1, \ldots, J + 1, \end{cases} \quad (2.25)$$

where J_* denotes the number of "rescaled parameters" in $h_n(\lambda)$, $J_* \leq q$, $\{d_{n,j} : n \geq 1\}$ is a non-decreasing sequence of constants that diverges to $\infty \, \forall j \leq J_*$, $m_j(\lambda) \in R \, \forall j = J_* + 1, \ldots, J$, and $m_{J+1}(\lambda) \in \mathcal{T}$ for some compact metric space $\mathcal{T}$.

In fact, often, one has $J_* = 1$, $m_j(\lambda) = \lambda_j$ and no term $m_{J+1}(\lambda)$ appears. If the CS is determined by a test statistic, as is usually the case, then the terms $d_{n,j}\lambda_j$ and $m_j(\lambda)$ are chosen so that the test statistic converges in distribution to some limit whenever $d_{n,j}\lambda_{n,j} \to h_j$ as $n \to \infty$ for $j = 1, \ldots, J_*$ and $m_j(\lambda_n) \to h_j$ as $n \to \infty$ for $j = J_* + 1, \ldots, J + 1$ for some sequence of parameters $\{\lambda_n \in \Lambda : n \geq 1\}$.

The scaling constants $\{d_{n,j} : n \geq 1\}$ often are $d_{n,j} = n^{1/2}$ or $d_{n,j} = n$ when a natural parametrization is employed.[6]

The function $h_n(\lambda)$ in (2.25) is designed to handle the case in which the pointwise asymptotic coverage probability of $\{CS_n : n \geq 1\}$ or the pointwise asymptotic distribution of a test statistic exhibits a discontinuity at any $\lambda \in \Lambda$ for which (a) $\lambda_j = 0$ for all $j \leq J_*$, or alternatively, (b) $\lambda_j = 0$ for some $j \leq J_*$, as in Cheng (2015). To see this, suppose $J_* = 1$, $\lambda \in \Lambda$ has $\lambda_1 = 0$, and $\lambda^\varepsilon \in \Lambda$ equals λ except that $\lambda_1^\varepsilon = \varepsilon > 0$. Then, under the (constant) sequence $\{\lambda : n \geq 1\}$, $h_{n,1}(\lambda) \to h_1 = 0$, whereas under the sequence $\{\lambda^\varepsilon : n \geq 1\}$, $h_{n,1}(\lambda^\varepsilon) \to h_1^\varepsilon = \infty$ no matter how close ε is to 0. Thus, if $h = \lim_{n\to\infty} h_n(\lambda)$ and $h^\varepsilon = \lim_{n\to\infty} h_n(\lambda^\varepsilon)$, then h does not equal h^ε because $h_1 = 0$ and $h_1^\varepsilon = \infty$ and h^ε does not depend on ε for $\varepsilon > 0$. Provided $CP^+(h) \neq CP^+(h^\varepsilon)$ and/or $CP^-(h) \neq CP^-(h^\varepsilon)$, there is a discontinuity in the pointwise asymptotic coverage probability of $\{CS_n : n \geq 1\}$ at λ.

The function $h_n(\lambda)$ in (2.25) can be reformulated to allow for discontinuities when $\lambda_j = \lambda_j^0$, rather than $\lambda_j = 0$, for all $j \leq J_*$ or for some $j \leq J_*$. To do so, one takes $h_{n,j}(\lambda) = d_{n,j}(\lambda_j - \lambda_j^0) \, \forall j \leq J_*$.

The function $h_n(\lambda)$ in (2.25) also can be reformulated to allow for discontinuities at J_* different values of a single parameter λ_k for some $k \leq q$, e.g., at the values in $\{\lambda_{k,1}^0, \ldots, \lambda_{k,J_*}^0\}$ (rather than a single discontinuity at multiple parameters $\{\lambda_1, \ldots, \lambda_k\}$). In this case, one takes $h_{n,j}(\lambda) = d_{n,j}(\lambda_k - \lambda_{k,j}^0)$ for $j = 1, \ldots, J_*$. The function $h_n(\lambda)$ can be reformulated to allow for multiple discontinuities at each parameter value in $\{\lambda_{k_1}, \ldots, \lambda_{k_L}\}$, where $k_\ell \leq q \, \forall \ell \leq L$, e.g., at the values in $\{\lambda_{k_\ell,1}^0, \ldots, \lambda_{k_\ell,J_{*,\ell}}^0\} \, \forall \ell \leq L$. In this case, one takes $h_{n,j}(\lambda) = d_{n,j}(\lambda_{k_\ell} - \lambda_{k_\ell,j-S_{\ell-1}}^0)$ for $j = S_{\ell-1} + 1, \ldots, S_\ell$ for $\ell = 1, \ldots, L$, where $S_\ell = \sum_{s=0}^{\ell} J_{*,s}$, $J_{*,0} = 0$ and $J_* = \sum_{s=0}^{L} J_{*,s}$.

A weaker and somewhat simpler assumption than Assumption B is the following.

Assumption B1. For any sequence $\{\lambda_n \in \Lambda : n \geq 1\}$ for which $h_n(\lambda_n) \to h \in H$, $CP_n(\lambda_n) \to [CP^-(h), CP^+(h)]$ for some $CP^-(h), CP^+(h) \in [0, 1]$.

The difference between Assumptions B and B1 is that Assumption B must hold for all subsequences $\{p_n\}$ for which "…" holds, whereas Assumption B1 only needs to hold for all sequences $\{n\}$ for which "…" holds. In practice, the same arguments that are used to verify Assumption B1 based on sequences usually also can be used to verify Assumption B for subsequences with very few changes. In both cases, one has to verify results under a triangular array framework. For example, the triangular array CLT for martingale differences given in Hall and Heyde (1980, Theorem 3.2, Corollary 3.1) and the triangular array empirical process results given in Pollard (1990, Theorem 10.6) can be employed.

Next, consider the following assumption.

Assumption B2. For any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ for which $h_{p_n}(\lambda_{p_n}) \to h \in H$, there exists a sequence $\{\lambda_n^* \in \Lambda : n \geq 1\}$ such that $h_n(\lambda_n^*) \to h \in H$ and $\lambda_{p_n}^* = \lambda_{p_n} \, \forall n \geq 1$.

---

[6] The scaling constants are arbitrary in the sense that if λ_j is reparametrized to be $(\lambda_j)^{\beta_j}$ for some $\beta_j > 0$, then $d_{n,j}$ becomes $d_{n,j}^{\beta}$. For example, in an AR(1) model with AR(1) parameter ρ, a natural parametrization is to take $\lambda_1 = 1 - \rho$, which leads to $d_{n,1} = n$. But, if one takes $\lambda_1 = (1-\rho)^\beta$, then $d_{n,1} = n^\beta$ for $\beta > 0$.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

If Assumption B2 holds, then Assumptions B and B1 are equivalent. In consequence, the following Lemma holds immediately.

Lemma 2.1. *Assumptions* B1 *and* B2 *imply Assumption* B.

Assumption B2 looks fairly innocuous, so one might consider imposing it and replacing Assumption B with Assumptions B1 and B2. However, in some cases, Assumption B2 can be difficult to verify or can require superfluous assumptions. In such cases, it is easier to verify Assumption B directly.

Under Assumption B2, H simplifies to

$$H = \{h \in (R \cup \{\pm\infty\})^J \times \mathcal{T} : h_n(\lambda_n) \to h \text{ for some sequence } \{\lambda_n \in \Lambda : n \geq 1\}\}. \tag{2.26}$$

Next, we provide a sufficient condition, Assumption B2*, for Assumption B2.

Assumption B2*. (i) For all $\lambda \in \Lambda$, $\lambda = (\lambda_1, \ldots, \lambda_q, \lambda_{q+1})'$, where $\lambda_j \in R \; \forall j \leq q$ and λ_{q+1} belongs to some metric space.
(ii) condition (ii) given in (2.25) holds.
(iii) $m_j(\lambda) (= m_j(\lambda_1, \ldots, \lambda_{q+1}))$ is continuous in $(\lambda_1, \ldots, \lambda_{J_*})$ uniformly over $\lambda \in \Lambda \; \forall j = J_* + 1, \ldots, J + 1$.[7]
(iv) The parameter space Λ satisfies: for some $\delta > 0$ and all $\lambda = (\lambda_1, \ldots, \lambda_{q+1})' \in \Lambda$, $(a_1\lambda_1, \ldots, a_{J_*}\lambda_{J_*}, \lambda_{J_*+1}, \ldots, \lambda_{q+1})' \in \Lambda \; \forall a_j \in (0, 1]$ if $|\lambda_j| \leq \delta$, where $a_j = 1$ if $|\lambda_j| > \delta$ for $j \leq J_*$.

The comments given above regarding (2.25) also apply to the function $h_n(\lambda)$ in Assumption B2*.

Lemma 2.2. *Assumption* B2* *implies Assumption* B2.

The proof of Lemma 2.2 is given in Section 2.3.

For simplicity, we combine Assumptions B and S, and B1 and S, as follows.

Assumption B*. For any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ for which $h_{p_n}(\lambda_{p_n}) \to h \in H$, $CP_{p_n}(\lambda_{p_n}) \to CP$ for some $CP \in [0, 1]$.

Assumption B1*. For any sequence $\{\lambda_n \in \Lambda : n \geq 1\}$ for which $h_n(\lambda_n) \to h \in H$, $CP_n(\lambda_n) \to CP$ for some $CP \in [0, 1]$.

The relationship among the assumptions is

$$\left.\begin{array}{l} B1^* \Rightarrow B1 \\ B2^* \Rightarrow B2 \end{array}\right\} \Rightarrow B, \quad B \Rightarrow A1 \;\&\; A2, \;\text{ and }\; \begin{array}{l} B1^* \Rightarrow S \\ B^* \Rightarrow S. \end{array} \\ B^* \Rightarrow B, \tag{2.27}$$

The results of the last two subsections are summarized as follows.

Corollary 2.1. *The confidence sets* $\{CS_n : n \geq 1\}$ *satisfy the following results.*
(a) *Under Assumption* B (*or* B1 *and* B2, *or B1 and* B2*), $AsySz \in [\inf_{h \in H} CP^-(h), \inf_{h \in H} CP^+(h)]$.
(b) *Under Assumptions* B, C1, *and* C2 (*or* B1, B2, C1, *and* C2, *or* B1, B2* , C1, *and* C2), $AsySz = \inf_{h \in H} CP^-(h) = \inf_{h \in H} CP^+(h)$ *and* $AsyMaxCP = \sup_{h \in H} CP^-(h) = \sup_{h \in H} CP^+(h)$.
(c) *Under Assumption* B* (*or* B1* *and* B2, *or* B1* *and* B2*), $AsySz = AsyMaxCP = CP$.

Comments. 1. Corollary 2.1(a) is used to establish the asymptotic size of CS's that are (i) not asymptotically similar and (ii) exhibit sufficient discontinuities in the asymptotic distribution functions of their test statistics under drifting sequences such that $\inf_{h \in H} CP^-(h) < \inf_{h \in H} CP^+(h)$. Property (ii) is not typical. Corollary 2.1(b) is used to establish the asymptotic size of CS's in the more common case where property (ii) does not hold and the CS's are not asymptotically similar. Corollary 2.1 (c) is used to establish the asymptotic size and asymptotic similarity of CS's that are asymptotically similar.
2. With the adjustments in Comments 4 and 5 to Theorem 2.1, the results of Corollary 2.1 also hold for tests.

2.3. Proofs for sufficient conditions

Proof of Theorem 2.2. First we show that Assumption B implies Assumption A1. Below we show Condition Sub: For any sequence $\{\lambda_n \in \Lambda : n \geq 1\}$ and any subsequence $\{w_n\}$ of $\{n\}$ there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that $h_{p_n}(\lambda_{p_n}) \to h$ for some $h \in H$. Given Condition Sub, we apply Assumption B to $\{p_n\}$ and $\{\lambda_{p_n}\}$ to get $CP_{p_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)]$ for some $h \in H$, which implies that Assumption A1 holds.

Now we establish Condition Sub. Let $\{w_n\}$ be some subsequence of $\{n\}$. Let $h_{w_n,j}(\lambda_{w_n})$ denote the jth component of $h_{w_n}(\lambda_{w_n})$ for $j = 1, \ldots, J+1$. Let $p_{1,n} = w_n \; \forall n \geq 1$. For $j = 1$, either (1) $\limsup_{n \to \infty} |h_{p_{j,n},j}(\lambda_{p_{j,n}})| < \infty$ or (2) $\limsup_{n \to \infty} |h_{p_{j,n},j}(\lambda_{p_{j,n}})| = \infty$. If (1) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$,

$$h_{p_{j+1,n},j}(\lambda_{p_{j+1,n}}) \to h_j \text{ for some } h_j \in R. \tag{2.28}$$

[7] That is, $\forall \varepsilon > 0$, $\exists \delta^* > 0$ such that $\forall \lambda, \lambda^* \in \Lambda$ with $\|(\lambda_1, \ldots, \lambda_{J_*}) - (\lambda_1^*, \ldots, \lambda_{J_*}^*)\| < \delta^*$ and $(\lambda_{J_*+1}, \ldots, \lambda_{q+1}) = (\lambda_{J_*+1}^*, \ldots, \lambda_{q+1}^*)$, $\rho(m_j(\lambda), m_j(\lambda^*)) < \varepsilon$, where $\rho(\cdot, \cdot)$ denotes Euclidean distance on R when $j \leq J$ and $\rho(\cdot, \cdot)$ denotes the metric on $\mathcal{T}$ when $j = J + 1$.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

If (2) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$,

$$h_{p_{j+1,n},j}(\lambda_{p_{j+1,n}}) \to h_j,\ \text{where } h_j = \infty \text{ or } -\infty. \tag{2.29}$$

Applying the same argument successively for $j = 2, \ldots, J$ yields a subsequence $\{p_n^*\} = \{p_{J+1,n}\}$ of $\{w_n\}$ for which $h_{p_n^*,j}(\lambda_{p_n^*}) \to h_j\ \forall j \le J$. Now, $\{h_{p_n^*,J+1}(\lambda_{p_n^*}) : n \ge 1\}$ is a sequence in the compact set $\mathcal{T}$. By compactness, there exists a subsequence $\{s_n : n \ge 1\}$ of $\{n\}$ such that $\{h_{p_{s_n}^*,J+1}(\lambda_{p_{s_n}^*}) : n \ge 1\}$ converges to an element of $\mathcal{T}$, call it h_{J+1}. The subsequence $\{p_n\} = \{p_{s_n}^*\}$ of $\{w_n\}$ is such that $h_{p_n}(\lambda_{p_n}) \to h = (h_1, \ldots, h_{J+1})' \in H$, which establishes Condition Sub.

Next, we show that Assumption B implies Assumption A2. Given any $h \in H$, by the definition of H in (2.24), there exists a subsequence $\{p_n\}$ and a sequence $\{\lambda_{p_n} \in \Lambda : n \ge 1\}$ such that $h_{p_n}(\lambda_{p_n}) \to h$. In consequence, by Assumption B, $CP_{p_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)]$ and Assumption A2 holds. □

Proof of Lemma 2.2. Let $\{p_n\}$ and $\{\lambda_{p_n}\}$ be as in Assumption B2. Then, $h_{p_n}(\lambda_{p_n}) \to h$, $d_{p_n,j}\lambda_{p_n,j} \to h_j\ \forall j \le J_*$, and $m_j(\lambda_{p_n}) \to h_j\ \forall j = J_* + 1, \ldots, J+1$ using Assumption B2*(ii).

Given $h \in H$ as in Assumption B2 and $\delta > 0$ as in Assumption B2*(iv), $\exists N < \infty$ such that $\forall n \ge N$, $|\lambda_{p_n,j}| < \delta\ \forall j \le J_*$ for which $|h_j| < \infty$. (This holds because $|h_j| < \infty$ and $d_{n,j} \to \infty$ imply that $\lambda_{p_n,j} \to 0$ as $n \to \infty\ \forall j \le J_*$.)

Define a new sequence $\{\lambda_s^* = (\lambda_{s,1}^*, \ldots, \lambda_{s,q}^*, \lambda_{s,q+1}^*)' : s \ge 1\}$ as follows: (i) $\forall s < p_N$, take λ_s^* to be an arbitrary element of Λ, (ii) $\forall s = p_n$ and $n \ge N$, define $\lambda_s^* = \lambda_{p_n} \in \Lambda$, and (iii) $\forall s \in (p_n, p_{n+1})$ and $n \ge N$, define

$$\lambda_{s,j}^* = \begin{cases} (d_{p_n,j}/d_{s,j})\lambda_{p_n,j} & \text{if } |h_j| < \infty \ \&\ j \le J_* \\ \lambda_{p_n,j} & \text{if } |h_j| = \infty \ \&\ j \le J_*,\ \text{ or if } j = J_* + 1, \ldots, J+1. \end{cases} \tag{2.30}$$

In case (iii), $d_{p_n,j}/d_{s,j} \in (0, 1]$ (for n large enough) by Assumption B2*(ii), and hence, using Assumption B2*(iv), we have $\lambda_s^* \in \Lambda$. Thus, $\lambda_s^* \in \Lambda\ \forall s \ge 1$.

For all $j \le J_*$ with $|h_j| < \infty$, we have $d_{s,j}\lambda_{s,j}^* = d_{p_n,j}\lambda_{p_n,j}\ \forall s \in [p_n, p_{n+1})$ with $s \ge p_N$, and $d_{p_n,j}\lambda_{p_n,j} \to h_j$ as $n \to \infty$ by the first paragraph of the proof. Hence, $h_{s,j}(\lambda_s^*) = d_{s,j}\lambda_{s,j}^* \to h_j$ as $s \to \infty$.

For all $j \le J_*$ with $h_j = \infty$, we have $d_{s,j}\lambda_{s,j}^* = d_{s,j}\lambda_{p_n,j} \ge d_{p_n,j}\lambda_{p_n,j}\ \forall s \in [p_n, p_{n+1})$ with $s \ge p_N$ and with s large enough that $\lambda_{s,j}^* > 0$ using the property that $d_{n,j}$ is non-decreasing in j by Assumption B2*(ii). We also have $d_{p_n,j}\lambda_{p_n,j} \to h_j = \infty$ as $n \to \infty$ by the first paragraph of the proof. Hence, $d_{s,j}\lambda_s^* \to h_j = \infty$ as $s \to \infty$. The argument for the case where $j \le J_*$ with $h_j = -\infty$ is analogous.

Next, we consider $j = J_* + 1, \ldots, J+1$. Define $\lambda_s^{**} = \lambda_{p_n}\ \forall s \in [p_n, p_{n+1})$ and all $n \ge N$ and $\lambda_s^{**} = \lambda_s^*\ \forall s \le p_N$. For $j = J_* + 1, \ldots, J+1$, $m_j(\lambda_{p_n}) \to h_j$ as $n \to \infty$ by the first paragraph of the proof, which implies that $m_j(\lambda_s^{**}) \to h_j$ as $s \to \infty$.

In the following, let ρ denote the Euclidean distance on R when $j \le J$ and let ρ denote the metric on $\mathcal{T}$ when $j = J+1$. Now, for $j = J_* + 1, \ldots, J+1$, if $|h_{j_1}| < \infty\ \forall j_1 \le J_*$, we have: $\forall s \in [p_n, p_{n+1})$ with $s \ge p_N$,

$$\begin{aligned} & \rho(m_j(\lambda_s^*), m_j(\lambda_s^{**})) \\ = \ & \rho(m_j((d_{p_n,1}/d_{s,1})\lambda_{p_n,1}, \ldots, (d_{p_n,J_*}/d_{s,J_*})\lambda_{p_n,J_*}, \lambda_{p_n,J_*+1}, \ldots, \lambda_{p_n,q+1}), \\ & m_j(\lambda_{p_n,1}, \ldots, \lambda_{p_n,J_*}, \lambda_{p_n,J_*+1}, \ldots, \lambda_{p_n,q+1})) \\ \to \ & 0 \text{ as } s \to \infty, \end{aligned} \tag{2.31}$$

where the convergence holds by Assumption B2*(iii) using the fact that $\lambda_{p_n,j_1} \to 0$ as $n \to \infty$ because $|h_{j_1}| < \infty$, $(d_{p_n,j_1}/d_{s,j_1}) \in [0, 1]\ \forall s \in [p_n, p_{n+1})$ by Assumption B2*(ii), and hence, $\sup_{s \in [p_n, p_{n+1})} |(d_{p_n,j_1}/d_{s,j_1})\lambda_{p_n,j_1}| \to 0$ as $n \to \infty$ and $\sup_{s \in [p_n, p_{n+1})} |(d_{p_n,j_1}/d_{s,j_1})\lambda_{p_n,j_1} - \lambda_{p_n,j_1}| \to 0$ as $n \to \infty\ \forall j_1 \le J_*$. Eq. (2.31) and $m_j(\lambda_s^{**}) \to h_j$ as $s \to \infty$ imply that $m_j(\lambda_s^*) \to h_j$ as $s \to \infty$, as desired. If $|h_{j_1}| = \infty$ for one or more $j_1 \le J_*$, then the corresponding elements of λ_s^* equal those of λ_s^{**} and the convergence in (2.31) still holds by Assumption B2*(iii). Hence, we conclude that for $j = J_* + 1, \ldots, J+1$, $m_j(\lambda_s^*) \to h_j$ as $s \to \infty$.

Replacing s by n, we conclude that $\{\lambda_n^* \in \Lambda : n \ge 1\}$ satisfies $h_n(\lambda_n^*) \to h \in H$ and $\lambda_{p_n}^* = \lambda_{p_n}\ \forall n \ge 1$ and so Assumption B2 holds. □

3. Conditional likelihood ratio test with weak instruments

In the following sections, we apply the theory above to a number of different examples. In this section, we consider a heteroskedasticity-robust version of Moreira's (2003) CLR test concerning the parameter on a scalar endogenous variable in the linear IV regression model. We show that this test (and corresponding CI) has asymptotic size equal to its nominal size and is asymptotically similar in a uniform sense with IV's that may be weak and errors that may be heteroskedastic.

Consider the linear IV regression model

$$\begin{aligned} y_1 &= y_2\theta + X\xi + u, \\ y_2 &= Z\pi + X\phi + v, \end{aligned} \tag{3.1}$$

where $y_1, y_2 \in R^n$ are vectors of endogenous variables, $X \in R^{n \times d_X}$ for $d_X \ge 0$ is a matrix of included exogenous variables, and $Z \in R^{n \times d_Z}$ for $d_Z \ge 1$ is a matrix of IV's. Denote by u_i and X_i the ith rows of u and X, respectively, written as column vectors (or scalars) and analogously for other random variables. Assume that $\{(X_i', Z_i', u_i, v_i)' : 1 \leq i \leq n\}$ are i.i.d. with distribution F. The vector $(\theta, \pi', \xi', \phi')' \in R^{1+d_Z+2d_X}$ consists of unknown parameters.

We are interested in testing the null hypothesis

$$H_0 : \theta = \theta_0 \tag{3.2}$$

against a two-sided alternative $H_1 : \theta \neq \theta_0$.

For any matrix B with n rows, let

$$B^{\perp} \equiv M_X B, \text{ where } M_A \equiv I_n - P_A, \quad P_A = A(A'A)^{-1}A' \tag{3.3}$$

for any full column rank matrix A, and I_n denotes the n-dimensional identity matrix. If no regressors X appear, then we set $M_X = I_n$. Note that from (3.1) we have $y_1^{\perp} = y_2^{\perp}\theta + u^{\perp}$ and $y_2^{\perp} = Z^{\perp}\pi + v^{\perp}$. Define

$$g_i(\theta) \equiv Z_i^{\perp}(y_{1i}^{\perp} - y_{2i}^{\perp}\theta) \text{ and } G_i \equiv -(\partial/\partial\theta')g_i(\theta) = Z_i^{\perp}y_{2i}^{\perp}. \tag{3.4}$$

Let

$$\overline{Z}_i \equiv (X_i', Z_i')' \text{ and } Z_i^* \equiv Z_i - E_F Z_i X_i'(E_F X_i X_i')^{-1} X_i, \tag{3.5}$$

where E_F denotes expectation under F. Let Z^* be the $n \times d_Z$ matrix with ith row $Z_i^{*\prime}$.

Now, we define the CLR test for $H_0 : \theta = \theta_0$. Let

$$g_i \equiv g_i(\theta_0), \quad \widehat{g} \equiv n^{-1}\sum_{i=1}^n g_i, \quad \widehat{G} \equiv n^{-1}\sum_{i=1}^n G_i, \quad \widehat{\Omega} \equiv n^{-1}\sum_{i=1}^n g_i g_i' - \widehat{g}\widehat{g}',$$

$$\widehat{\Psi} \equiv \widehat{\Sigma} - \widehat{\Gamma}\widehat{\Omega}^{-1}\widehat{\Gamma}, \quad \widehat{\Sigma} \equiv n^{-1}\sum_{i=1}^n Z_i^{\perp}Z_i^{\perp\prime}\widehat{v}_i^2, \quad \widehat{\Gamma} \equiv n^{-1}\sum_{i=1}^n Z_i^{\perp}Z_i^{\perp\prime}\widehat{u}_i\widehat{v}_i,$$

$$\widehat{u} \equiv M_X(y_1 - y_2\theta_0) \ (= u^{\perp}), \text{ and } \widehat{v} \equiv M_{\overline{Z}}y_2 \ (\neq v^{\perp}). \tag{3.6}$$

For notational convenience, subscripts n are omitted.[8]

We define the Anderson and Rubin (1949) statistic and the Lagrange Multiplier statistic of Kleibergen (2002) and Moreira (2009), generalized to allow for heteroskedasticity, as follows:

$$AR \equiv n\widehat{g}'\widehat{\Omega}^{-1}\widehat{g} \text{ and } LM \equiv n\widehat{g}'\widehat{\Omega}^{-1/2}P_{\widehat{\Omega}^{-1/2}\widehat{D}}\widehat{\Omega}^{-1/2}\widehat{g}, \text{ where}$$

$$\widehat{D} \equiv \widehat{G} - n^{-1}\sum_{i=1}^n (G_i - \widehat{G})g_i'\widehat{\Omega}^{-1}\widehat{g}. \tag{3.7}$$

Note that $\widehat{D}$ equals (minus) the derivative with respect to θ of the moment conditions $n^{-1}\sum_{i=1}^n g_i(\theta)$ with an adjustment to make the latter asymptotically independent of the moment conditions $\widehat{g}$.

We define a Wald statistic for testing $\pi = 0$ as follows:

$$W \equiv n\widehat{D}'\widehat{\Psi}^{-1}\widehat{D}. \tag{3.8}$$

A small value of W indicates that the IV's are weak.

The heteroskedasticity-robust CLR test statistic is

$$CLR \equiv \frac{1}{2}\left(AR - W + \sqrt{(AR - W)^2 + 4LM \cdot W}\right). \tag{3.9}$$

The CLR statistic has the property that for W large it is approximately equal to the LM statistic LM.[9]

The critical value of the CLR test is $c(1-\alpha, W)$. Here, $c(1-\alpha, w)$ is the $(1-\alpha)$-quantile of the distribution of

$$clr(w) \equiv \frac{1}{2}\left(\chi_1^2 + \chi_{d_Z-1}^2 - w + \sqrt{(\chi_1^2 + \chi_{d_Z-1}^2 - w)^2 + 4\chi_1^2 w}\right), \tag{3.10}$$

where χ_1^2 and $\chi_{d_Z-1}^2$ are independent chi-square random variables with 1 and $d_Z - 1$ degrees of freedom, respectively. Critical values are given in Moreira (2003).[10] The nominal $100(1-\alpha)\%$ CLR CI for θ is the set of all values θ_0 for which the CLR test fails to reject $H_0 : \theta = \theta_0$. Fast computation of this CI can be carried out using the algorithm of Mikusheva and Poi (2006). The function $c(1-\alpha, w)$ is decreasing in w and equals $\chi_{d_Z,1-\alpha}^2$ and $\chi_{1,1-\alpha}^2$ when $w = 0$ and ∞, respectively, see Moreira (2003), where $\chi_{m,1-\alpha}^2$ denotes the $(1-\alpha)$ -quantile of a chi-square distribution with m degrees of freedom.

8 The centering of $\widehat{\Omega}$ using $\widehat{g}\widehat{g}'$ has no effect asymptotically under the null and under local alternatives, but it does have an effect under non-local alternatives.

9 To see this requires some calculations, see the proof of Lemma 3.1 in Appendix.

10 More extensive tables of critical values are given in the Supplemental Material to Andrews et al. (2006), which is available on the Econometric Society website.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

The CLR test rejects the null hypothesis $H_0: \theta = \theta_0$ if

$$CLR > c(1-\alpha, W). \tag{3.11}$$

With homoskedastic normal errors, the CLR test of Moreira (2003) has some approximate asymptotic optimality properties within classes of equivariant similar and non-similar tests under both weak and strong IV's, see Andrews et al. (2006, 2008, 2019) and Mikusheva (2010). Under homoskedasticity, the heteroskedasticity-robust CLR test defined here has the same null and local alternative asymptotic properties as the homoskedastic CLR test of Moreira (2003). Hence, with homoskedastic normal errors, it possesses the same approximate asymptotic optimality properties. By the results established below, it also has correct asymptotic size and is asymptotically similar under both homoskedasticity and heteroskedasticity (for any strength of the IV's and errors that need not be normal). Mikusheva (2010) establishes that a "homoskedastic" CLR test has correct asymptotic size under homoskedasticity of the errors. This test does not have correct asymptotic size under conditional heteroskedasticity, which is the most relevant situation in practice. Mikusheva (2010) establishes uniformity using an a.s. representation argument, which is a different approach from that taken in this paper. Mikusheva's approach does not yield uniformity over a class of distributions of the errors, exogenous variables, and instruments, which is obtained by the approach taken in this paper.

The heteroskedasticity-robust test in (3.11) is similar to the SR-CQLR test in Andrews and Guggenberger (2019) when the latter is specialized to the linear IV model with a single rhs endogenous variable, which is considered here.[11] The test is not exactly the same because the latter test is designed to be robust to singularity of the variance matrix of the moments.

Next, we define the parameter space for the null distributions that generate the data. Define

$$Var_F\begin{pmatrix} Z_i^* u_i \\ Z_i^* v_i \end{pmatrix} = \begin{bmatrix} \Omega_F & \Gamma_F \\ \Gamma_F & \Sigma_F \end{bmatrix} \equiv \begin{bmatrix} E_F Z_i^* Z_i^{*\prime} u_i^2 & E_F Z_i^* Z_i^{*\prime} u_i v_i \\ E_F Z_i^* Z_i^{*\prime} u_i v_i & E_F Z_i^* Z_i^{*\prime} v_i^2 \end{bmatrix} \text{ and } \Psi_F \equiv \Sigma_F - \Gamma_F \Omega_F^{-1} \Gamma_F'. \tag{3.12}$$

Under the null hypothesis, the distribution of the data is determined by $\lambda = (\lambda_1, \lambda_2, \lambda_{3F}, \lambda_4, \lambda_{5F})$, where

$$\lambda_1 = \|\pi\|, \lambda_2 = \pi/\|\pi\|, \lambda_{3F} = (E_F Z_i^* Z_i^{*\prime}, \Omega_F, \Sigma_F, \Gamma_F), \lambda_4 = (\xi, \phi), \tag{3.13}$$

and λ_{5F} equals F, the distribution of $(u_i, v_i, X_i', Z_i')'$.[12] By definition, $\pi/\|\pi\|$ equals $1_{d_Z}/d_Z^{1/2}$ if $\|\pi\| = 0$, where $\|\cdot\|$ denotes the Euclidean norm. As defined, λ completely determines the distribution of the observations. As is well-known, vectors π close to the origin lead to weak IV's. Hence, $\lambda_1 = \|\pi\|$ measures the strength of the IV's. We define λ_{3F} as in (3.13) because the asymptotic distribution of the components of the CLR test statistic under a sequence of distributions $\{F_n\}$ depends on the limits of the elements of λ_{3F_n}, see (A.76)–(A.80) in Appendix.

The parameter space Λ of null distributions is

$$\begin{aligned} \Lambda \equiv \{\lambda = (\lambda_1, \lambda_2, \lambda_{3F}, \lambda_4, \lambda_{5F}) : (\pi, \xi, \phi) \in R^{d_Z + 2d_X}, \\ \lambda_{5F}\ (= F) \text{ satisfies } E_F \overline{Z}_i(u_i, v_i) = 0,\ \ E_F\|\overline{Z}_i\|^{k_1}|u_i|^{k_2}|v_i|^{k_3} \le M \\ \text{for all } k_1, k_2, k_3 \ge 0,\ \ k_1 + k_2 + k_3 \le 4 + \delta,\ \ k_2 + k_3 \le 2 + \delta, \\ \lambda_{\min}(A) \ge \delta \text{ for } A = \Psi_F, \Sigma_F, \Omega_F, E_F \overline{Z}_i \overline{Z}_i', E_F Z_i^* Z_i^{*\prime}\} \end{aligned} \tag{3.14}$$

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

When applying the results of Section 2, we let

$$h_n(\lambda) = (n^{1/2}\lambda_1, \lambda_2, \lambda_{3F}) \tag{3.15}$$

and we take H as defined in (2.24) with no $(J+1)^{th}$ component present.

Assumption B* holds with $RP = \alpha\ \forall h \in H$ by the following Lemma.[13]

Lemma 3.1. *The asymptotic null rejection probability of the nominal level α CLR test equals α under all subsequences $\{p_n\}$ and all sequences $\{\lambda_{p_n} \in \Lambda : n \ge 1\}$ for which $h_{p_n}(\lambda_{p_n}) \to h \in H$.*

Given that Assumption B* holds, Corollary 2.1(c) implies that the asymptotic size of the CLR test equals its nominal size and the CLR test is asymptotically similar (in a uniform sense). Correct asymptotic size of the CLR CI, rather than the CLR test, requires uniformity over θ_0. This holds automatically because the finite-sample distribution of the CLR test for testing $H_0: \theta = \theta_0$ when θ_0 is the true value is invariant to θ_0.

Furthermore, the proof of Lemma 3.1 shows that the AR and LM tests, which reject $H_0: \theta = \theta_0$ when $AR(\theta_0) > \chi^2_{d_Z, 1-\alpha}$ and $LM(\theta_0) > \chi^2_{1,1-\alpha}$, respectively, satisfy Assumption B* with $RP = \alpha$. Hence, these tests also have asymptotic size equal to their nominal size and are asymptotically similar (in a uniform sense). However, the CLR test has better power than these tests in an overall sense.

---

11 The reason is that the $\lambda_{\min}(nQ_n(\theta))$ quantity that appears in the definition of the $QLR_n(\theta)$ test statistic in (5.7) of Andrews and Guggenberger (2019) can be given a closed-form expression because $Q_n(\theta)$ is a 2×2 matrix in this case, which leads to an expression for $QLR_n(\theta)$ like that for CLR in (3.9).

12 For notational simplicity, we let λ (and some other quantities below) be concatenations of vectors and matrices.

13 Here, RP is the testing analogue of CP. See Comment 4 to Theorem 2.1.

4. Grid bootstrap CI in an AR(k) model

Hansen (1999) proposes a grid bootstrap CI for parameters in an AR(k) model. Using the results in Section 2, we show this grid bootstrap CI has correct asymptotic size and is asymptotically similar in a uniform sense. The parameter space over which the uniform result is established is specified below. Mikusheva (2007) also demonstrates the uniform asymptotic validity and similarity of the grid bootstrap CI. Compared to Mikusheva (2007), our results include uniformity over the innovation distribution, which is an infinite-dimensional nuisance parameter, whereas hers do not, see Mikusheva (2007, p. 1341). Our approach does not use almost sure representations, which are employed in Mikusheva (2007). It just uses asymptotic coverage probabilities under drifting subsequences of parameters.

We focus on the grid bootstrap CI for ρ_1 in the Augmented Dickey–Fuller (ADF) representation of the AR(k) model:

$$Y_t = \mu_0 + \mu_1 t + \rho_1 Y_{t-1} + \rho_2 \Delta Y_{t-1} + \cdots + \rho_k \Delta Y_{t-k+1} + U_t, \tag{4.1}$$

where $\rho_1 \in [-1+\varepsilon, 1]$ for some $\varepsilon > 0$, U_t is i.i.d. with unknown distribution F, and $E_F U_t = 0$.[14] The time series $\{Y_t : t \geq 1\}$ is initialized at some fixed value $(Y_0, \ldots, Y_{1-k})'$.[15] Let $\rho = (\rho_1, \ldots, \rho_k)'$. Define a lag polynomial $a(L;\rho) \equiv 1 - \rho_1 L - \sum_{j=2}^{k} \rho_j L^{j-1}(1-L)$ and factorize it as $a(L;\rho) = \Pi_{j=1}^{k}(1-\gamma_j(\rho)L)$, where $|\gamma_1(\rho)| \leq \cdots \leq |\gamma_k(\rho)|$. Note that $1-\rho_1 = \Pi_{j=1}^{k}(1-\gamma_j(\rho))$. We have $\rho_1 = 1$ if and only if $\gamma_k(\rho) = 1$. We construct a CI for ρ_1 under the assumption that $|\gamma_k(\rho)| \leq 1$ and $|\gamma_{k-1}(\rho)| \leq 1-\delta$ for some $\delta > 0$.

In this example, $\lambda = (\rho, F)$. The parameter space for λ is

$$\begin{aligned} \Lambda \equiv \{&(\rho, F) : \rho_1 \in [-1+\varepsilon, 1],\ \ \rho \in \Omega^*,\ \ F \in \mathcal{F}^*\},\ \text{where} \\ &\Omega^* \text{ is some compact subset of } \Omega, \\ &\Omega \equiv \{\rho : |\gamma_k(\rho)| \leq 1, |\gamma_{k-1}(\rho)| \leq 1-\delta\}, \\ &\mathcal{F}^* \text{ is some compact subset of } \mathcal{F} \text{ wrt to Mallow's metric } d_{2r},\ \text{and} \\ &\mathcal{F} \equiv \{F : E_F U_t = 0 \text{ and } E_F U_t^4 \leq M\} \end{aligned} \tag{4.2}$$

for some $\varepsilon > 0$, $\delta > 0$, $M < \infty$, and $r \geq 1$. Mallows (1972) metric d_{2r} also is used in Hansen (1999).

The t statistic is used to construct the grid bootstrap CI. By definition, $t_n(\rho_1) = (\widehat{\rho}_1 - \rho_1)/\widehat{\sigma}_1$, where $\widehat{\rho}_1$ denotes the least squares (LS) estimator of ρ_1 and $\widehat{\sigma}_1$ denotes the standard error estimator of $\widehat{\rho}_1$.

The grid bootstrap CI for ρ_1 is constructed as follows. For $\lambda \in \Lambda$, let $(\widehat{\rho}_2(\rho_1), \ldots, \widehat{\rho}_k(\rho_1))'$ denote the constrained LS estimator of $(\rho_2, \ldots, \rho_k)'$ given ρ_1 for the model in (4.1). Let $\widehat{F}$ denote the empirical distribution of the residuals from the unconstrained LS estimator of ρ based on (4.1), as in Hansen (1999). Bootstrap samples $\{Y_t(\rho_1) : t \leq n\}$ are simulated using (4.1) with $(\rho_1, \widehat{\rho}_2(\rho_1), \ldots, \widehat{\rho}_k(\rho_1), \widehat{F})'$ in place of $(\rho_1, \rho_2, \ldots, \rho_k, F)'$, with 0 in place of μ_0, and with some fixed starting values $(Y_0, \ldots, Y_{1-k})'$, such as $(0, \ldots, 0)'$.[16] Bootstrap t statistics are constructed using the bootstrap samples. Let $q_n^*(\alpha|\rho_1)$ denote the α quantile of the empirical distribution of the bootstrap t statistics. The grid bootstrap CI for ρ_1 is

$$C_{g,n} = \{\rho_1 \in [-1+\varepsilon, 1] : q_n^*(\alpha/2|\rho_1) \leq t_n(\rho_1) \leq q_n^*(1-\alpha/2|\rho_1)\}, \tag{4.3}$$

where $1-\alpha$ is the nominal coverage probability of the CI.

To show that the grid bootstrap CI $C_{g,n}$ has asymptotic size equal to $1-\alpha$, we consider sequences of true parameters $\{\lambda_n = (\rho_n, F_n) \in \Lambda : n \geq 1\}$ such that $n(1-\rho_{1,n}) \to h_1 \in [0,\infty]$ and $\lambda_n \to \lambda_0 = (\rho_0, F_0) \in \Lambda$, where $\rho_n \equiv (\rho_{1,n}, \ldots, \rho_{k,n})'$ and $\rho_0 \equiv (\rho_{1,0}, \ldots, \rho_{k,0})'$. Define

$$\begin{aligned} h_n(\lambda) &= (n(1-\rho_1), \lambda')' \text{ and} \\ H &= \{(h_1, \lambda_0')' \in [0,\infty] \times \Lambda:\ n(1-\rho_{1,n}) \to h_1 \\ &\quad\ \text{and } \lambda_n \to \lambda_0 \text{ for some } \{\lambda_n \in \Lambda : n \geq 1\}\}. \end{aligned} \tag{4.4}$$

Lemma 4.1. *For all sequences $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ for which $h_{p_n}(\lambda_{p_n}) \to (h_1, \lambda_0) \in H$, $CP_{p_n}(\lambda_{p_n}) \to 1-\alpha$ for C_{g,p_n} defined in (4.3) with n replaced by p_n.*

Comment. The proof of Lemma 4.1 uses results in Hansen (1999), Giraitis and Phillips (2006), and Mikusheva (2007).[17]

---

[14] The results given below could be extended to martingale difference innovations with constant conditional variances without much difficulty.

[15] The model can be extended to allow for random starting values, for example, along the lines of Andrews and Guggenberger (2012). More specifically, from (4.1), Y_t can be written as $Y_t = \mu_0^* + \mu_1^* t + Y_t^*$ and $Y_t^* = \rho_1 Y_{t-1}^* + \rho_2 \Delta Y_{t-1}^* + \cdots + \rho_k \Delta Y_{t-k+1}^* + U_t$ for some μ_0^* and μ_1^*. Let $y_0^* = (Y_0^*, \ldots, Y_{1-k}^*)$ denote the starting value for $\{Y_t^* : t \geq 1\}$. When $\rho_1 < 1$, the distribution of y_0^* can be taken to be the distribution that yields strict stationarity for $\{Y_t^* : t \geq 1\}$. When $\rho_1 = 1$, y_0^* can be taken to be arbitrary. With these starting values, the asymptotic distribution of the t statistic under near unit-root parameter values changes, but the asymptotic size and similarity results given below do not change.

[16] The bootstrap starting values can be different from those for the original sample.

[17] The results from Mikusheva (2007) that are employed in the proof are not related to uniformity issues. They are an extension from an AR(1) model to an AR(k) model of an L^2 convergence result for the least squares covariance matrix estimator and a martingale difference central limit theorem for the score, which are established in Giraitis and Phillips (2006).

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

In order to establish a uniform result, Lemma 4.1 covers (i) the stationary case, i.e., $h_1 = \infty$ and $\rho_{1,0} \neq 1$, (ii) the near stationary case, i.e., $h_1 = \infty$ and $\rho_{1,0} = 1$, (iii) the near unit-root case, i.e., $h_1 \in R$ and $\rho_{1,0} = 1$, and (iv) the unit-root case, i.e., $h_1 = 0$ and $\rho_{1,0} = 1$. In the proof of Lemma 4.1, we show that $t_n(\rho_1) \to_d N(0, 1)$ in cases (i) and (ii), even though the rate of convergence of the LS estimator of ρ_1 is non-standard (faster than $n^{1/2}$) in case (ii). In cases (iii) and (iv), $t_n(\rho_1) \Rightarrow (\int_0^1 W_c dW)/(\int_0^1 W_c^2)^{1/2}$, where $W_c(r) = \int_0^r \exp(r-s)dW(s)$, $W(s)$ is standard Brownian motion, and $c = h_1/(1-\sum_{j=2}^k \rho_{j,0})$.

Lemma 4.1 implies that Assumption B* holds for the grid bootstrap CI. By Corollary 2.1(c), the asymptotic size of the grid bootstrap CI equals its nominal size and the grid bootstrap CI is asymptotically similar (in a uniform sense).

5. Quasi-likelihood ratio confidence intervals in nonlinear regression

In this example, we consider the asymptotic properties of standard quasi-likelihood ratio-based CI's in a nonlinear regression model. We determine the *AsySz* of such CI's and find that they are not necessarily equal to their nominal size. We also determine the degree of asymptotic non-similarity of the CI's, which is defined by *AsyMaxCP* − *AsySz*. We make use of results given in Andrews and Cheng (2012b, Appendix E) concerning the asymptotic properties of the LS estimator in the nonlinear regression model and general asymptotic properties of QLR test statistics under drifting sequences of distributions. But note that Andrews and Cheng (2012b) do not consider QLR-based tests or CI's in the nonlinear regression model.

5.1. Nonlinear regression model

The model is

$$Y_i = \beta \cdot h(X_i, \pi) + Z_i'\zeta + U_i \text{ for } i = 1, \ldots, n, \tag{5.1}$$

where $Y_i \in R$, $X_i \in R^{d_X}$, and $Z_i \in R^{d_Z}$ are observed i.i.d. random variables or vectors, $U_i \in R$ is an unobserved homoskedastic i.i.d. error term, and $h(X_i, \pi) \in R$ is a function that is known up to the parameter $\pi \in R$. When the true value of β is zero, (5.1) becomes a linear model and π is not identified. This non-regularity is the source of the asymptotic size problem.

We analyze QLR-based CI's for the scalars β and π.

The Gaussian quasi-likelihood function leads to the nonlinear LS estimator of the parameter $\theta = (\beta, \zeta', \pi)'$. The LS sample criterion function is

$$Q_n(\theta) \equiv n^{-1}\sum_{i=1}^n U_i^2(\theta)/2, \text{ where } U_i(\theta) \equiv Y_i - \beta h(X_i, \pi) - Z_i'\zeta. \tag{5.2}$$

Note that when $\beta = 0$, the residual $U_i(\theta)$ and the criterion function $Q_n(\theta)$ do not depend on π.

The (unrestricted) LS estimator $\widehat{\theta}_n$ of θ minimizes $Q_n(\theta)$ over $\theta \in \Theta$. The optimization parameter space Θ takes the form

$$\Theta = \mathcal{B} \times \mathcal{Z} \times \Pi, \text{ where } \mathcal{B} = [-b_1, b_2] \subset R, \tag{5.3}$$

$\mathcal{Z}(\subset R^{d_\zeta})$ is compact, and $\Pi(\subset R)$ is compact.

The random variables $\left\{(X_i', Z_i', U_i)' : i = 1, \ldots, n\right\}$ are i.i.d. with distribution ϕ. The support of X_i (for all possible true distributions of X_i) is contained in a set $\mathcal{X}$. We assume that $h(x, \pi)$ is twice continuously differentiable wrt π, $\forall \pi \in \Pi$, $\forall x \in \mathcal{X}$. Let $h_\pi(x, \pi) \in R$ and $h_{\pi\pi}(x, \pi) \in R$ denote the first-order and second-order partial derivatives of $h(x, \pi)$ wrt π.

The parameter space for the true value of θ is

$$\Theta^* = \mathcal{B}^* \times \mathcal{Z}^* \times \Pi^*, \text{ where } \mathcal{B}^* = [-b_1^*, b_2^*] \subset R, \tag{5.4}$$

$b_1^* \geq 0$, $b_2^* \geq 0$, b_1^* and b_2^* are not both equal to 0, $\mathcal{Z}^*(\subset R^{d_\zeta})$ is compact, and $\Pi^*(\subset R)$ is compact.[18] Let Φ^* be a space of distributions of (X_i, Z_i, U_i) that is a compact metric space with some metric that induces weak convergence of these distributions (such as the supremum norm of the distribution functions). The parameter space for the true value of ϕ is

$$\begin{aligned}
\Phi^{**} \equiv \{&\phi \in \Phi^* : E_\phi(U_i|X_i, Z_i) = 0 \text{ a.s.}, \quad E_\phi(U_i^2|X_i, Z_i) = \sigma^2 > 0 \text{ a.s.},\\
&E_\phi\left(\sup_{\pi\in\Pi}|h(X_i, \pi)|^{4+\varepsilon} + \sup_{\pi\in\Pi}|h_\pi(X_i, \pi)|^{4+\varepsilon} + \sup_{\pi\in\Pi}|h_{\pi\pi}(X_i, \pi)|^{2+\varepsilon}\right) \leq C,\\
&|h_{\pi\pi}(X_i, \pi_1) - h_{\pi\pi}(X_i, \pi_2)| \leq M(X_i)|\pi_1 - \pi_2| \;\; \forall \pi_1, \pi_2 \in \Pi \text{ for some function}\\
&M(X_i), \quad E_\phi M(X_i)^{2+\varepsilon} \leq C, \quad E_\phi|U_i|^{4+\varepsilon} \leq C, \quad E_\phi\|Z_i\|^{4+\varepsilon} \leq C,\\
&P_\phi(a'(h(X_i, \pi_1), h(X_i, \pi_2), Z_i) = 0) < 1, \quad \forall \pi_1, \pi_2 \in \Pi \text{ with } \pi_1 \neq \pi_2, \quad \forall a \in R^{d_\zeta + 2}
\end{aligned}$$

[18] We allow the optimization parameter space Θ and the "true parameter space" Θ^*, which includes the true parameter by definition, to be different to avoid boundary issues. Provided $\Theta^* \subset int(\Theta)$, as is assumed below, boundary problems do not arise.

$$\text{with } a \neq 0, \quad \lambda_{\min}(E_\phi(h(X_i,\pi), Z_i')'(h(X_i,\pi), Z_i')) \geq \varepsilon \ \ \forall \pi \in \Pi, \text{ and}$$
$$\lambda_{\min}(E_\phi d_i(\pi)\, d_i(\pi)') \geq \varepsilon \ \ \forall \pi \in \Pi\} \tag{5.5}$$

for some constants $C < \infty$ and $\varepsilon > 0$, and by definition $d_i(\pi) = (h(X_i, \pi), Z_i, h_\pi(X_i, \pi))'$. The moment conditions in Φ^{**} are used to ensure the uniform convergence of various sample averages. The other conditions are used for the identification of β and ζ and the identification of π when $\beta \neq 0$.

We assume that the optimization parameter space Θ is chosen such that $b_1 > b_1^*$, $b_2 > b_2^*$, $\mathcal{Z}^* \in int(\mathcal{Z})$, and $\mathcal{B}^* \in int(\mathcal{B})$. This ensures that the true parameter cannot be on the boundary of the optimization parameter space.

5.2. Confidence intervals

We consider CI's for β and π.[19] The CI's are obtained by inverting tests. For the CI for β, we consider tests of the null hypothesis

$$H_0 : r(\theta) = v, \text{ where } r(\theta) = \beta. \tag{5.6}$$

For the CI for π, the function $r(\theta)$ is $r(\theta) = \pi$.

For $v \in r(\Theta)$, we define a restricted estimator $\widetilde{\theta}_n(v)$ of θ subject to the restriction that $r(\theta) = v$. By definition,

$$\widetilde{\theta}_n(v) \in \Theta, \quad r(\widetilde{\theta}_n(v)) = v, \text{ and } Q_n(\widetilde{\theta}_n(v)) = \inf_{\theta \in \Theta : r(\theta) = v} Q_n(\theta) + o(n^{-1}). \tag{5.7}$$

For testing $H_0 : r(\theta) = v$, the QLR test statistic is

$$QLR_n(v) \equiv 2n(Q_n(\widetilde{\theta}_n(v)) - Q_n(\widehat{\theta}_n))/\widehat{\sigma}_n^2, \text{ where}$$
$$\widehat{\sigma}_n^2 \equiv \widehat{\sigma}_n^2(\widehat{\theta}_n) \text{ and } \widehat{\sigma}_n^2(\theta) \equiv n^{-1} \sum_{i=1}^n U_i^2(\theta). \tag{5.8}$$

The critical value used with the standard QLR test statistic for testing a scalar restriction is the $1-\alpha$ quantile of the χ_1^2 distribution, which we denote by $\chi_{1,1-\alpha}^2$. This choice is based on the pointwise asymptotic distribution of the QLR statistic when $\beta \neq 0$.

The nominal level $1-\alpha$ QLR CS for $r(\theta) = \beta$ or $r(\theta) = \pi$ is

$$CS_n^{QLR} = \{v \in r(\Theta) : QLR_n(v) \leq \chi_{1,1-\alpha}^2\}. \tag{5.9}$$

5.3. Asymptotic results

Under sequences $\{(\theta_n, \phi_n) : n \geq 1\}$ such that $\theta_n \in \Theta^*$, $\phi_n \in \Phi^{**}$, $(\theta_n, \phi_n) \to (\theta_0, \phi_0)$, $\theta_0 = (\beta_0, \zeta_0', \pi_0)'$, $\beta_0 = 0$, and $n^{1/2}\beta_n \to b \in R$ (which implies that $|b| < \infty$), we have the following result:

$$QLR_n \to_d LR_\infty(b, \pi_0, \phi_0) = 2(\inf_{\pi \in \Pi_{r,0}} \xi_r(\pi; b, \pi_0, \phi_0) - \inf_{\pi \in \Pi} \xi(\pi; b, \pi_0, \phi_0))/\sigma_0^2, \tag{5.10}$$

where $\Pi_{r,0} = \Pi$ if $r(\theta) = \beta$, $\Pi_{r,0} = \pi_0$ if $r(\theta) = \pi$, σ_0^2 denotes the variance of U_i under ϕ_0, and the stochastic processes $\{\xi(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$ and $\{\xi_r(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$ are defined below in Section 5.5.[20]

We assume that the distribution function of $LR_\infty(b, \pi_0, \phi_0)$ is continuous at $\chi_{1,1-\alpha}^2$ $\forall b \in R$, $\forall \pi_0 \in \Pi^*$, $\forall \phi_0 \in \Phi^{**}$.[21] It is difficult to provide primitive sufficient conditions for this assumption to hold. However, given the Gaussianity of the processes underlying $LR_\infty(b, \pi_0, \phi_0)$, it typically holds. For completeness, we provide results both when this condition holds and when it fails.

Next, under sequences $\{(\theta_n, \phi_n) : n \geq 1\}$ such that $\theta_n \in \Theta^*$, $\phi_n \in \Phi^{**}$, $(\theta_n, \phi_n) \to (\theta_0, \phi_0)$, $n^{1/2}|\beta_n| \to \infty$, and $\beta_n/|\beta_n| \to \omega_0 \in \{-1, 1\}$, we have the result:

$$QLR_n \to_d \chi_1^2. \tag{5.11}$$

The results in (5.10) and (5.11) are proved in the Appendix using results in Andrews and Cheng (2012b).

19 CI's for elements of ζ and some nonlinear functions of β, π, and ζ can be obtained from the results given here by verifying Assumptions RQ1–RQ3 in the Appendix.

20 The random quantity $\xi(\pi; b, \pi_0, \phi_0)$ is the limiting distribution under $\{(\theta_n, \phi_n) : n \geq 1\}$ (that satisfies the specified conditions) of the concentrated criterion function $Q_n(\widehat{\theta}_n(\pi))$ after suitable centering and scaling, where $\widehat{\theta}_n(\pi)$ minimizes $Q_n(\theta)$ over Θ for given $\pi \in \Pi$. Analogously, $\xi_r(\pi; b, \pi_0, \phi_0)$ is the limiting distribution under $\{(\theta_n, \phi_n) : n \geq 1\}$ of the restricted concentrated criterion function $Q_n(\widetilde{\theta}_n(v, \pi))$ after suitable centering and scaling, where $\widetilde{\theta}_n(v, \pi)$ minimizes $Q_n(\theta)$ over Θ subject to the restriction $r(\theta) = v$ for given $\pi \in \Pi_{r,0}$.

21 This assumption is stronger than needed, but it is simple. It is sufficient that the distribution function of $LR_\infty(b, \pi_0, \phi_0)$ is continuous at $\chi_{1,1-\alpha}^2$ for (b, π_0, ϕ_0) equal to some (b_L, π_L, ϕ_L) and (b_U, π_U, ϕ_U) in $R \times \Pi^* \times \Phi^{**}$ for which $P(LR_\infty(b_L, \pi_L, \phi_L) < \chi_{1,1-\alpha}^2) = \inf_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} P(LR_\infty(b, \pi_0, \phi_0) < \chi_{1,1-\alpha}^2)$ and $P(LR_\infty(b_U, \pi_U, \phi_U) \leq \chi_{1,1-\alpha}^2) = \sup_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} P(LR_\infty(b, \pi_0, \phi_0) \leq \chi_{1,1-\alpha}^2)$.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

Now, we apply the results of Corollary 2.1(b) above with $\lambda = (|\beta|, \beta/|\beta|, \zeta', \pi, \phi)'$, $h_n(\lambda) = (n^{1/2}|\beta|, |\beta|, \beta/|\beta|, \zeta', \pi, \phi)'$, where by definition $\beta/|\beta| = 1$ if $\beta = 0$, and $h = (b, |\beta_0|, \beta_0/|\beta_0|, \zeta_0', \pi_0, \phi_0)'$. We verify Assumptions B1 and B2*, C1, and C2. Assumption B1 holds by (5.10) and (5.11) with

$$\begin{aligned}
H &= \{(b, |\beta_0|, \beta_0/|\beta_0|, \zeta_0', \pi_0, \phi_0)' \in [-\infty, \infty] \times \mathcal{B}^* \times \mathcal{Z}^* \times \Pi^* \times \Phi^{**} : \\
&\quad sgn(b) = sgn(\beta_0) \text{ if } sgn(\beta_0) \neq 0\}, \\
CP^-(h) &= CP^+(h) = CP(b, \pi_0, \phi_0) \text{ when } |b| < \infty, \text{ where} \\
CP(b, \pi_0, \phi_0) &= P\left(LR_\infty(b, \pi_0, \phi_0) \leq \chi^2_{1,1-\alpha}\right). \qquad (5.12)
\end{aligned}$$

When $|b| = \infty$, Assumption B1 holds with

$$CP^-(h) = CP^+(h) = P\left(S \leq \chi^2_{1,1-\alpha}\right) = 1 - \alpha, \qquad (5.13)$$

where S is a random variable with χ^2_1 distribution. Hence, Assumptions C1 and C2 also hold using the condition on the distribution function (df) of $LR_\infty(b, \pi_0, \phi_0)$.

Assumption B2*(i) holds with $(\lambda_1, \ldots, \lambda_q)' = (|\beta|, \beta/|\beta|, \zeta', \pi)'$ and $\lambda_{q+1} = \phi$. Assumption B2*(ii) holds with $h_n(\lambda)$ as above, $r = 1$, $d_{n,j} = n^{1/2}$, $J = 4 + d_\zeta$, $m_j(\lambda) = \lambda_{j-1}$ for $j = 2, \ldots, J$, $m_{J+1}(\lambda) = \phi$, $\Lambda = \{\lambda = (|\beta|, \beta/|\beta|, \zeta', \pi, \phi)' : \theta = (\beta, \zeta', \pi)' \in \Theta^*, \phi \in \Phi^{**}\}$, and $\mathcal{T} = \Phi^* \supset \Phi^{**}$, where Φ^* is a compact metric space by assumption. Assumption B2*(iii) holds immediately given the form of $m_j(\lambda)$ for $j = 2, \ldots, J+1$. Assumption B2*(iv) holds because given any $\lambda = (|\beta|, \beta/|\beta|, \zeta', \pi, \phi)' \in \Lambda$, $(a|\beta|, \beta/|\beta|, \zeta', \pi, \phi)' \in \Lambda$ for all $a \in (0, 1]$ by the form of Λ and $\Theta^* = \mathcal{B}^* \times \mathcal{Z}^* \times \Pi^*$, where $\mathcal{B}^* = [-b_1^*, b_2^*] \subset R$ with $b_1^* \geq 0$, $b_2^* \geq 0$, and b_1^* and b_2^* not both equal to 0.

Hence, by Corollary 2.1(b), we have

$$\begin{aligned}
AsySz &= \min\{\inf_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} CP(b, \pi_0, \phi_0), 1 - \alpha\} \text{ and} \\
AsyMaxCP &= \max\{\sup_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} CP(b, \pi_0, \phi_0), 1 - \alpha\}. \qquad (5.14)
\end{aligned}$$

Typically, $AsySz < 1 - \alpha$ and the QLR CI for β or π does not have correct asymptotic size. Below we provide a numerical example for a particular choice of $h(x, \pi)$.

Using the general approach in Andrews and Cheng (2012a), one can construct data-dependent critical values (that are used in place of the fixed critical value $\chi^2_{1,1-\alpha}$) that yield QLR-based CI's for β and π with $AsySz$ equal to their nominal size. For brevity, we do not provide details here.

If the continuity condition on the df of $LR_\infty(b, \pi_0, \phi_0)$ does not hold, then Assumptions C1 and C2 do not necessarily hold. In this case, instead of (5.14), using Corollary 2.1(a), we have

$$\begin{aligned}
AsySz &\in \Big[\min\{\inf_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} CP^-(b, \pi_0, \phi_0), 1 - \alpha\}, \\
&\qquad \min\{\inf_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} CP(b, \pi_0, \phi_0), 1 - \alpha\}\Big], \text{ and} \\
AsyMaxCP &\in \Big[\max\{\sup_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} CP^-(b, \pi_0, \phi_0), 1 - \alpha\}, \\
&\qquad \max\{\sup_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} CP(b, \pi_0, \phi_0), 1 - \alpha\}\Big], \text{ where} \\
CP^-(b, \pi_0, \phi_0) &= P\left(LR_\infty(b, \pi_0, \phi_0) < \chi^2_{1,1-\alpha}\right). \qquad (5.15)
\end{aligned}$$

5.4. Numerical results

Here we compute $AsySz$ and $AsyMaxCP$ in (5.14) for two choices of the nonlinear regression function $h(x, \pi)$. We compute these quantities for a single distribution ϕ (i.e., for the case where Φ^{**} contains a single element). The model is as in (5.1) with $Z_i = (1, Z_i^*)' \in R^2$.

The two nonlinear functions considered are:

(i) a Box–Cox function $h(x, \pi) = (x^\pi - 1)/\pi$ and
(ii) a logistic function $h(x, \pi) = x(1 + \exp(-(x - \pi)))^{-1}$. (5.16)

In both cases, $\{(Z_i^*, X_i, U_i) : i = 1, \ldots, n\}$ are i.i.d. and U_i is independent of (Z_i^*, X_i).

When the Box–Cox function is employed, $Z_i^* \sim N(0, 1)$, $X_i = |X_i^*|$ with $X_i^* \sim N(3, 1)$, $Corr(Z_i^*, X_i^*) = 0.5$, and $U_i \sim N(0, 0.5^2)$. The true values of ζ_0 and ζ_1 are -2 and 2, respectively. The true values of π considered are $\{1.50, 1.75, \ldots, 3.5\}$. The optimization space Π for π is $[1, 4]$.

Table 1
Asymptotic coverage probabilities of nominal 95% standard QLR CI's for β and π in the nonlinear regression model.

Box–Cox function								
	π_0	1.50	2.00	2.50	3.00	3.50	*AsySz*	*AsyMaxCP*
β	min over b	0.918	0.918	0.918	0.918	0.918	0.918	
	max over b	0.987	0.992	0.989	0.986	0.974		0.992
π	min over b	0.950	0.950	0.950	0.950	0.950	0.950	
	max over b	0.997	0.999	0.999	0.998	0.995		0.999
Logistic function								
	π_0	4.5	4.7	5.0	5.2	5.5	*AsySz*	*AsyMaxCP*
β	min over b	0.744	0.744	0.744	0.744	0.744	0.744	
	max over b	0.953	0.951	0.950	0.951	0.951		0.953
π	min over b	0.868	0.869	0.869	0.869	0.870	0.868	
	max over b	0.950	0.953	0.951	0.951	0.950		0.953

When the logistic function is employed, $Z_i^* \sim N(5,1)$, $X_i = Z_i^*$, $U_i \sim N(0,1)$. The true values of ζ_0 and ζ_1 are -5 and 1, respectively. The true values of π considered are $\{4.5, 4.6, \ldots, 5.5\}$. The optimization space Π for π is $[4,6]$, where the lower and upper bounds are approximately the 15% and 85% quantiles of X_i.

In both cases, the discrete values of b for which computations are made run from 0 to 20 (although only values from 0 to 5 are reported in Fig. 1), with a grid of 0.1 for b between 0 and 5, a grid of 0.2 for b between 5 and 10, and a grid of 1 for b between 10 and 20. The number of simulation repetitions is 50, 000.

Table 1 reports *AsySz* and *AsyMaxCP* defined in (5.14) and the minimum and maximum of the asymptotic coverage probabilities over b, for several values of π_0. To calculate *AsySz* and *AsyMaxCP*, the values of π_0 considered are the true values of π given above. Fig. 1 plots the asymptotic coverage probability as a function of b.

Table 1 and Fig. 1 show that the properties of QLR CI depend greatly on the context. The QLR CI for π in the Box–Cox model has correct *AsySz*, whereas the QLR CI's for β in the Box–Cox model and β and π in the logistic model have incorrect asymptotic sizes. Furthermore, the asymptotic sizes of the QLR CI's for β and π in the logistic model are quite low, being .75 and .87, respectively. In the Box–Cox model, there is over-coverage for almost all parameter configurations. In contrast, in the logistic model, there is under-coverage for all parameter configurations. In both models, there is a substantial degree of asymptotic nonsimilarity. The values of *AsyMaxCP* − *AsySz* for β and π are .074 and .049, respectively, in the Box–Cox model and .209 and .085, respectively, in the logistic model.

5.5. Quantities in the asymptotic distribution

Now, we define $\xi(\pi; b, \pi_0, \phi_0)$ and $\xi_r(\pi; b, \pi_0, \phi_0)$, which appear in (5.10). The stochastic process $\{\xi(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$ depends on the following quantities:

$$\begin{aligned} H(\pi;\phi_0) &\equiv E_{\phi_0} d_{\psi,i}(\pi) d_{\psi,i}(\pi)', \\ K(\pi;\pi_0,\phi_0) &\equiv -E_{\phi_0} h(X_i,\pi_0) d_{\psi,i}(\pi), \text{ and} \\ \Omega(\pi_1,\pi_2;\phi_0) &\equiv \sigma_0^2 E_{\phi_0} d_{\psi,i}(\pi_1) d_{\psi,i}(\pi_2)', \text{ where} \\ d_{\psi,i}(\pi) &\equiv (h(X_i,\pi), Z_i')'. \end{aligned} \tag{5.17}$$

Let $G(\pi;\phi_0)$ denote a mean 0_{1+d_ζ} Gaussian process with covariance kernel $\Omega(\pi_1,\pi_2;\phi_0)$. The process $\{\xi(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$ is a "weighted non-central chi-square" process defined by

$$\xi(\pi; b, \pi_0, \phi_0) \equiv -\frac{1}{2}\left(G(\pi;\phi_0) + K(\pi;\pi_0,\phi_0)b\right)' H^{-1}(\pi;\phi_0)\left(G(\pi;\phi_0) + K(\pi;\pi_0,\phi_0)b\right). \tag{5.18}$$

Given the definition of Φ^{**}, $\{\xi(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$ has bounded continuous sample paths a.s.

Next, we define the "restricted" process $\{\xi_r(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$. Define the Gaussian process $\{\tau(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$ by

$$\tau(\pi; b, \pi_0, \phi_0) \equiv -H^{-1}(\pi;\phi_0)(G(\pi;\phi_0) + K(\pi;\pi_0,\phi_0)b) - (b, 0_{d_\zeta}')', \tag{5.19}$$

where $(b, 0_{d_\zeta}')' \in R^{1+d_\zeta}$.[22]

[22] The process $\{\tau(\pi; b, \pi_0, \gamma_0) : \pi \in \Pi\}$ arises in the formula for $\xi_r(\pi; b, \pi_0, \gamma_0)$ below because the asymptotic distribution of $n^{1/2}(\widehat{\beta}_n - \beta_n, \widehat{\zeta}_n' - \zeta_n')'$ under $\{(\theta_n, \phi_n) : n \geq 1\}$ such that $\theta_n \in \Theta^*$, $\phi_n \in \Phi^*(\theta_n)$, $(\theta_n, \phi_n) \to (\theta_0, \phi_0)$, $\beta_0 = 0$, and $n^{1/2}\beta_n \to b \in R^{d_\beta}$ is the distribution of $\tau(\pi^*(b, \pi_0, \phi_0); b, \pi_0, \phi_0)$, where $\pi^*(b, \pi_0, \phi_0) = \arg\min_{\pi \in \Pi} \xi(\pi; b, \pi_0, \phi_0)$.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

ECONOM: 4875

ARTICLE IN PRESS

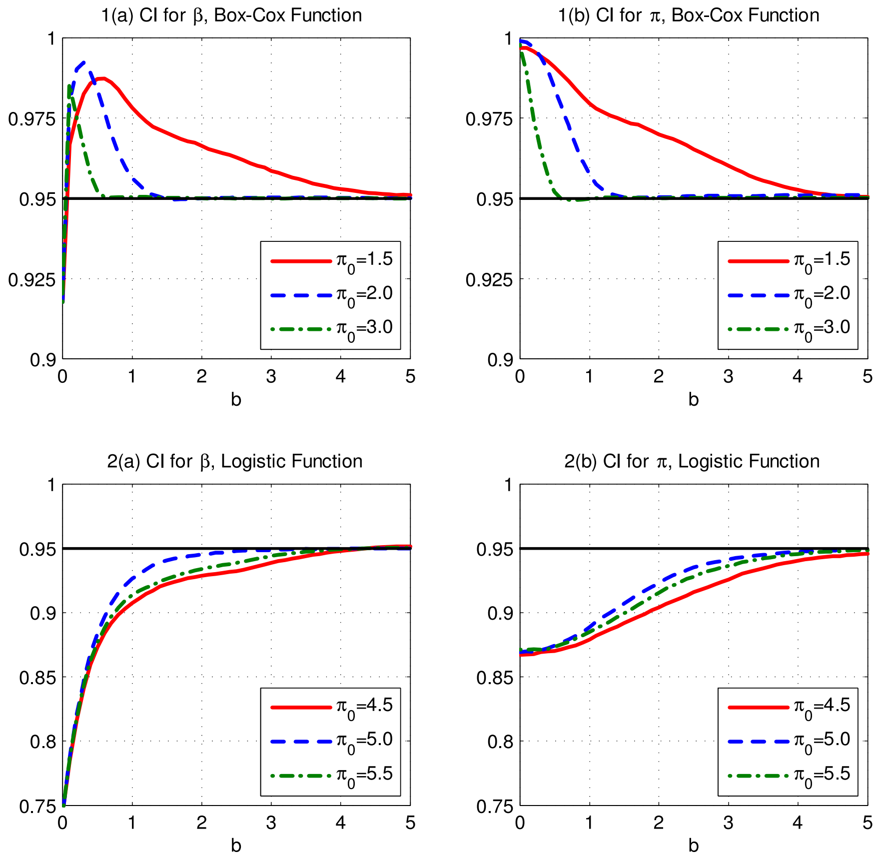

Fig. 1. Asymptotic Coverage Probabilities of Standard QLR CI's for β and π in the Nonlinear Regression Model.

The process $\{\xi_r(\pi; b, \pi_0, \phi_0) : \pi \in \Pi\}$ is defined by

$$\xi_r(\pi; b, \pi_0, \phi_0) \equiv \xi(\pi; b, \pi_0, \phi_0) + \frac{1}{2}\tau(\pi; b, \pi_0, \phi_0)' P_\psi(\pi; \phi_0)' H(\pi; \phi_0) P_\psi(\pi; \phi_0) \tau(\pi; b, \pi_0, \phi_0), \text{ where}$$
$$P_\psi(\pi; \phi_0) \equiv H^{-1}(\pi; \phi_0) e_1 \left(e_1' H^{-1}(\pi; \phi_0) e_1\right)^{-1} e_1'. \tag{5.20}$$

If $r(\theta) = \beta$, then $e_1 = (1, 0_{d_\zeta}')'$. If $r(\theta) = \pi$, then $e_1 = 0_{1+d_\zeta}$ and hence $\xi_r(\pi; b, \pi_0, \phi_0) = \xi(\pi; b, \pi_0, \phi_0)$. The $(1+d_\zeta) \times (1+d_\zeta)$-matrix $P_\psi(\pi; \phi_0)$ is an oblique projection matrix that projects onto the space spanned by e_1.

Appendix

This Appendix contains proofs of (i) results stated for the AR(1) running example, (ii) Lemma 3.1 concerning the conditional likelihood ratio test in the linear IV regression model considered in Section 3 of the paper, (iii) Lemma 4.1 concerning the grid bootstrap CI in an AR(k) model given in Section 4 of the paper, and (iv) Eqs. (5.10) and (5.11) for the nonlinear regression model considered in Section 5 of the paper.

A.1. Proofs for AR(1) running example

Here we derive the limiting distribution of the t-statistic under drifting sequences of parameters for the AR(1) example. The lemma is a simplified, self-contained version of Theorem 1 in Andrews and Guggenberger (2012) to the case where innovations of the AR(1) process are i.i.d. rather than stationary and strong mixing.

Lemma A.1. *Under drifting sequences $\{\lambda_n = (\rho_n, F_n) \in \Lambda : n \geq 1\}$ of* AR *parameters ρ_n and distributions F_n that satisfy*

$$n(1-\rho_n) \to h \in H = [0, \infty],$$

we have

$$T_n(\rho_n) \to_d J_h \text{ for } h \in H,$$

where $T_n(\rho_n)$, Λ, *and* J_h *are defined in* (2.10), (2.12), *and around* (2.16), *respectively.*

The proof of Lemma A.1 uses four lemmas stated below.

To simplify notation, we omit the subscript F_n on expectations. In integral expressions below, we often leave out the lower and upper limits zero and one, the argument r, and dr to simplify notation when there is no danger of confusion. For example, $\int_0^1 I_h(r)^2 dr$ is typically written as $\int I_h^2$. By " $\Rightarrow$ " we denote weak convergence of a stochastic process as $n \to \infty$.

Define $h_n^* \geq 0$ by $\rho_n = \exp(-h_n^*/n)$. By recursive substitution, we have

$$Y_{n,i}^* = \widetilde{Y}_{n,i} + \exp(-h_n^* i/n) Y_{n,0}^*, \text{ where}$$
$$\widetilde{Y}_{n,i} = \sum_{j=1}^{i} \exp(-h_n^*(i-j)/n) U_{n,j}. \tag{A.1}$$

For $\rho_n < 1$, we consider the following stationary start-up condition.

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

For the next four lemmas suppose Assumption STAT holds for n large and that asymptotics are derived under drifting sequences $\{\lambda_n = (\rho_n, F_n) \in \Lambda : n \geq 1\}$ of AR parameters ρ_n and distributions F_n. The first two lemmas deal with the case of $h \in [0, \infty)$. The last two deal with the case of $h = \infty$. Their proofs are given at the end of this subsection.

Lemma A.2. *Suppose* $\rho_n \in (-1, 1)$, $\rho_n = 1 - h_n/n$, *and* $h_n \to h \in [0, \infty)$ *as* $n \to \infty$. *Then, we have*

$$(2h_n/n)^{1/2} Y_{n,0}^* / \sigma_{U,n} \to_d Z \sim N(0, 1).$$

The next lemma is a special case of Lemma 1 in Phillips (1987) in the sense that we assume innovations to be i.i.d. while Phillips (1987) allows for non-i.i.d. strong mixing innovations. On the other hand, while Phillips (1987) assumes a startup of order $O_p(1)$ we allow a stationary startup. While Phillips (1987) assumes $\rho_n = 1 - h/n$ we allow for $\rho_n = 1 - h_n/n$ where $h_n \to h \in (0, \infty)$.

Lemma A.3. *Suppose* $\rho_n \in (-1, 1]$, $\rho_n = 1 - h_n/n$, *and* $h_n \to h \in (0, \infty)$. *Then, the following results* (a)–(k) *hold jointly for* $\sigma_{U,n}^2 = E_{F_n} U_{n,i}^2$:

(a) $n^{-1/2} Y_{n,[nr]}^* / \sigma_{U,n} \Rightarrow I_h^*(r)$, *where* $[a]$ *denotes the integer part of* a,
(b) $n^{-1} \sum_{i=1}^n U_{n,i} \to_p 0$,
(c) $n^{-1} \sum_{i=1}^n U_{n,i}^2 / \sigma_{U,n}^2 \to_p 1$,
(d) $n^{-1/2} \sum_{i=1}^n U_{n,i} / \sigma_{U,n} \to_d \int dW(r)$,
(e) $n^{-3/2} \sum_{i=1}^n Y_{n,i-1}^* / \sigma_{U,n} \to_d \int I_h^*(r) dr$,
(f) $n^{-1} \sum_{i=1}^n Y_{n,i-1}^* U_{n,i} / \sigma_{U,n}^2 \to_d \int I_h^*(r) dW(r)$,
(g) $n^{-2} \sum_{i=1}^n Y_{n,i-1}^{*2} / \sigma_{U,n}^2 \to_d \int I_h^{*2}(r) dr$, *and*
(h) *when* $h = 0$, *the parts of the lemma above hold with* $Y_{n,\cdot}^*$ *and* $I_h^*(\cdot)$ *replaced by* $\widetilde{Y}_{n,\cdot}$ *and* $I_h(\cdot)$, *respectively.*

Lemma A.4. *Suppose* $n(1 - \rho_n) \to \infty$ *and* $\rho_n \to 1$. *Then,*
(a) $EY_{n,0}^{*2} = (1 - \rho_n^2)^{-1} \sigma_{U,n}^2$,
(b) $n^{-1}(1 - \rho_n)^{1/2} X_1' X_2 = o_p(1)$,
(c) $(1 - \rho_n^2) n^{-1} X_1' X_1 / \sigma_{U,n}^2 \to_p 1$, *and*
(d) $(1 - \rho_n^2)^{1/2} n^{-1} U' X_1 = o_p(1)$.

Lemma A.5. *Suppose* $n(1 - \rho_n) \to \infty$ *and* $\rho_n \to 1$. *Let* $\mathcal{G}_{n,i}$ *the sequence of* σ*-fields (for* $i = \cdots, 1, 2, \ldots$ *for* $n \geq 1$*) generated by the* $U_{n,j}$ *for* $j \leq i$. *Then, we have*
(a) $\sum_{i=1}^n E(\zeta_{n,i}^2 1(|\zeta_{n,i}| > \delta) | \mathcal{G}_{n,i-1}) \to_p 0$ *for any* $\delta > 0$, *where*

$$\zeta_{n,i} \equiv n^{-1/2} \frac{Y_{i-1}^* U_{n,i}}{(1 - \rho_n^2)^{-1/2} \sigma_{U,n}^2}, \text{ and}$$

(b)

$$E\left(\sum_{i=1}^n E(\zeta_{n,i}^2 | \mathcal{G}_{n,i-1}) - 1\right)^2 \to 0.$$

The following lemma verifies Assumptions A1 and A2 in in Section 2.

Lemma A.6. *Assumptions* A1 *and* A2 *hold for the* AR(1) *example introduced in Example 1 of Section* 2 *with* $\lambda_n = (\rho_n, F_n)$, Λ *defined in* (2.12), $h \in H = [0, \infty]$, *and* $CP^-(h) = CP^+(h) = 1 - \alpha$.

The following lemma is used in the proof of Lemma A.6.

Lemma A.7. *For all sequences* $\{h_n \geq 0\}_{n \geq 1}$ *for which* $h_n \to h \in [0, \infty]$, $J_{h_n} \to_d J_h$.

To simplify notation further, we often leave out the subscript n here. For example, instead of ρ_n, $\sigma^2_{U,n} = EU^2_{n,i}$, $Y^*_{n,i}$, and $U_{n,i}$, we write ρ, σ^2_U, Y^*_i, and U_i. We do not drop n from h_n because h_n and h are different quantities.

Proof of Lemma A.1. We split the proof into three parts. Parts (i), (ii), and (iii) deal with the cases $h > 0$, $h = 0$, and $h = \infty$, which are the near unit root case, the (essentially) unit root case, and the stationary case, respectively.

Part (i). We can write

$$\begin{aligned} n(\widehat{\rho}_n - \rho) &= \left(n^{-2}X_1'M_{X_2}X_1\right)^{-1} n^{-1}X_1'M_{X_2}U \text{ and} \\ n\widehat{\sigma}^2_n &= \left(n^{-2}X_1'M_{X_2}X_1\right)^{-1} n^{-1}U'M_{X_2}M_{M_{X_2}X_1}M_{X_2}U \\ &= \left(n^{-2}X_1'M_{X_2}X_1\right)^{-1} [n^{-1}U'M_{X_2}U - (n^{-3/2}U'M_{X_2}X_1)^2(n^{-2}X_1'M_{X_2}X_1)^{-1}]. \end{aligned} \tag{A.2}$$

We next derive the limits of each of the components in the expressions in (A.2). Note that

$$\begin{aligned} & n^{-2}X_1'M_{X_2}X_1/\sigma^2_U \\ &= n^{-2}\sum_{i=1}^n \left(Y^*_{i-1} - \left(\sum_{j=1}^n Y^*_{j-1}/n\right)\right)^2 /\sigma^2_U \\ &= n^{-2}\sum_{i=1}^n Y^{*2}_{i-1}/\sigma^2_U - \left(n^{-3/2}\sum_{j=1}^n Y^*_{j-1}\right)^2 /\sigma^2_U \\ &\to_d \int I^{*2}_h - \left(\int I^*_h\right)^2 = \int I^{*2}_{D,h}, \end{aligned} \tag{A.3}$$

where the first two equalities hold by the definitions and straightforward algebra, and the convergence holds by Lemma A.3(e) and (g).

Similarly, we have

$$\begin{aligned} & n^{-1}X_1'M_{X_2}U/\sigma^2_U \\ &= n^{-1}\sum_{i=1}^n \left(Y^*_{i-1} - \left(\sum_{j=1}^n Y^*_{j-1}/n\right)\right) U_i/\sigma^2_U \\ &= n^{-1}\sum_{i=1}^n Y^*_{i-1}U_i/\sigma^2_U - \left(n^{-3/2}\sum_{j=1}^n Y^*_{j-1}/\sigma_U\right) n^{-1/2}\sum_{i=1}^n U_i/\sigma_U \\ &\to_d \int I^*_h dW - \int I^*_h \int dW = \int I^*_{D,h}dW, \end{aligned} \tag{A.4}$$

where the first two equalities hold by definitions and some algebra, and the convergence holds by Lemma A.3(d)–(f). Furthermore, by (A.3), (A.4), and Lemma A.3(b)–(c) it follows that

$$n^{-1}U'M_{X_2}U/\sigma^2_U - (n^{-3/2}U'M_{X_2}X_1/\sigma^2_U)^2(n^{-2}X_1'M_{X_2}X_1/\sigma^2_U)^{-1} \to_p 1. \tag{A.5}$$

Putting (A.3), (A.4), and (A.5) together gives $T^*_n(\rho) \to_d \left(\int I^{*2}_{D,h}\right)^{-1/2} \int I^*_{D,h}dW$, which completes the proof when $h > 0$.

Part (ii). In this case, (A.2)–(A.4) hold except that the convergence results in (A.3) and (A.4) only hold with Y^*_{i-1} replaced by $\widetilde{Y}_{i-1}$ because Lemma A.3 (h) only applies to random variables based on a zero initial condition when $h = 0$. Hence, we need to show that the difference between the second line of (A.3) with Y^*_{i-1} appearing and with $\widetilde{Y}_{i-1}$ appearing is $o_p(1)$ and that analogous results hold for (A.4).

For $h = 0$, by a mean value expansion, we have

$$\begin{aligned} & \max_{0 \leq j \leq 2n} |1 - \rho^j| = \max_{0 \leq j \leq 2n} |1 - \exp(-h^*_n j/n)| = \max_{0 \leq j \leq 2n} |1 - (1 - h^*_n j \exp(m_j)/n)| \\ & \leq 2h^*_n \max_{0 \leq j \leq 2n} |\exp(m_j)| = O(h^*_n), \end{aligned} \tag{A.6}$$

for an m_j such that $0 \le |m_j| \le h_n^* j/n \le 2h_n^* \to 0$, where h_n^* is defined just above (A.1).

Using the decomposition in (A.1), we have $Y_{i-1}^* = \widetilde{Y}_{i-1} + \rho^{i-1} Y_0^*$. To show the desired result for (A.3), the second line of (A.3) can be written as

$$
\begin{aligned}
& n^{-2}\sum_{i=1}^{n}\left(Y_{i-1}^* - \left(\sum_{j=1}^{n} Y_{j-1}^*/n\right)\right)^2/\sigma_U^2 \\
&= n^{-2}\sum_{i=1}^{n}\left(\widetilde{Y}_{i-1} + \rho^{i-1}Y_0^* - \left(\sum_{j=1}^{n} \widetilde{Y}_{j-1} + \rho^{j-1}Y_0^*\right)/n\right)^2/\sigma_U^2 \\
&= n^{-2}\sum_{i=1}^{n}\left(\widetilde{Y}_{i-1} - \left(\sum_{j=1}^{n} \widetilde{Y}_{j-1}/n\right) + O_p(h_n^* Y_0^*)\right)^2/\sigma_U^2 \qquad \text{(A.7)} \\
&= n^{-2}\sum_{i=1}^{n}\left(\widetilde{Y}_{i-1} - \left(\sum_{j=1}^{n} \widetilde{Y}_{j-1}/n\right)\right)^2/\sigma_U^2 + O_p(n^{-1/2}h_n^* Y_0^*),
\end{aligned}
$$

where the second equality holds because $\rho^{i-1} = 1 + O(h_n^*)$ uniformly in $i \le n$ by (A.6), and the third equality holds using Lemmas A.2 and A.3(e). Next, Lemma A.2 and $h_n^*/h_n \to 1$ (which is established at the beginning of the proof of Lemma A.2) show that $n^{-1/2}h_n^* Y_0^* = O_p(h_n^{*1/2}) = o_p(1)$. This completes the proof of the desired result for (A.3) when $h = 0$. The proof for (A.4) is similar and omitted.

Part (iii). It remains to consider the case where $h = \infty$. It is enough to consider the two cases $\rho \to \rho^* < 1$ and $\rho \to 1$.

First, assume $\rho \to 1$. By assumption, $n(1-\rho) \to \infty$. Define $a_n = n^{1/2}(1-\rho^2)^{-1/2}$. We first prove $a_n(\widehat{\rho}_n - \rho) \to_d N(0, 1)$. Note that

$$
a_n(\widehat{\rho}_n - \rho) = \left(n^{-1}\frac{X_1' M_{X_2} X_1}{(1-\rho^2)^{-1}\sigma_U^2}\right)^{-1} \frac{n^{-1/2} X_1' M_{X_2} U}{((1-\rho^2)^{-1}\sigma_U^4)^{1/2}} \equiv \nu_n \xi_n, \qquad \text{(A.8)}
$$

where ν_n and ξ_n have been implicitly defined as the first and second factor, respectively. By Lemma A.4(b) and (c),

$$
\frac{n^{-1}X_1'X_1}{(1-\rho^2)^{-1}\sigma_U^2} \to_p 1 \text{ and } \frac{n^{-1}X_1' P_{X_2} X_1}{(1-\rho^2)^{-1}\sigma_U^2} \to_p 0, \qquad \text{(A.9)}
$$

which implies $\nu_n \to_p 1$. We next show $\xi_n \to_d N(0, 1)$. Define

$$
\zeta_i \equiv n^{-1/2}\frac{Y_{i-1}^* U_i}{(1-\rho^2)^{-1/2}\sigma_U^2}. \qquad \text{(A.10)}
$$

It is enough to show that

$$
\frac{n^{-1/2}X_1' P_{X_2} U}{((1-\rho^2)^{-1}\sigma_U^4)^{1/2}} \to_p 0 \text{ and } \sum_{i=1}^{n} \zeta_i \to_d N(0, 1). \qquad \text{(A.11)}
$$

To show the first result, note that $n^{-1/2}X_2'U = n^{-1/2}\sum_{i=1}^{n} U_i = O_p(1)$ by a central limit theorem (CLT) for a triangular array of zero mean random variables U_i for which $E|U_i|^4 \le M$. The first result then follows from $(n^{-1}X_2'X_2)^{-1} = 1$ and Lemma A.4(b). To show the second result, by a CLT for a triangular array of martingale differences as in Corollary 3.1 in Hall and Heyde (1980), it is sufficient to establish the Lindeberg condition $\sum_{i=1}^{n} E(\zeta_i^2 1(|\zeta_i| > \delta)|\mathcal{G}_{i-1}) \to_p 0$ for any $\delta > 0$ and $\sum_{i=1}^{n} E(\zeta_i^2|\mathcal{G}_{i-1}) \to_p 1$. Lemma A.5(a) shows the former and Lemma A.5(b) together with Markov's inequality implies the latter.

We next show that $(1-\rho^2)^{-1/2}\widehat{\sigma}_n \to_p 1$, where

$$
(1-\rho^2)^{-1}\widehat{\sigma}_n^2 = \nu_n \frac{n^{-1}}{\sigma_U^2}[U' M_{X_2} U - (U' M_{X_2} X_1)^2 (X_1' M_{X_2} X_1)^{-1}]. \qquad \text{(A.12)}
$$

As established above $\nu_n \to_p 1$ and by a weak law of large numbers (WLLNs) $\frac{n^{-1}}{\sigma_U^2} U' M_{X_2} U \to_p 1$. Finally,

$$
\frac{n^{-1}}{\sigma_U^2}(U' M_{X_2} X_1)^2 (X_1' M_{X_2} X_1)^{-1} = O(\frac{n^{-2}}{(1-\rho^2)^{-1}}(U' M_{X_2} X_1)^2)\nu_n = o_p(1), \qquad \text{(A.13)}
$$

where for the last equality we use Lemma A.4(b) and (d).

Putting the statements above together, establishes $T_n^*(\rho) \to_d N(0, 1)$ for the case $\rho \to 1$ and $n(1-\rho) \to \infty$.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

Finally, consider the case $\rho \to \rho^* < 1$. This case is proven by applications of appropriate CLTs and WLLNs. For example, to show $v_n \to_p 1$ note that

$$\begin{aligned} & n^{-1}\frac{X_1' M_{X_2} X_1}{(1-\rho^2)^{-1}\sigma_U^2} - 1 \\ &= \frac{\sum_{i=1}^n Y_{i-1}^{*2}/n - (\sum_{i=1}^n Y_{i-1}^*/n)^2 - (1-\rho^2)^{-1}\sigma_U^2}{(1-\rho^2)^{-1}\sigma_U^2} \\ &= \frac{\sum_{i=1}^n \left(\sum_{j=0}^\infty \rho^j U_{i-1-j}\right)^2/n - (\sum_{i=1}^n \sum_{j=0}^\infty \rho^j U_{i-1-j}/n)^2 - (1-\rho^2)^{-1}\sigma_U^2}{(1-\rho^2)^{-1}\sigma_U^2}. \end{aligned} \tag{A.14}$$

The term $\sum_{i=1}^n \sum_{j=0}^\infty \rho^j U_{i-1-j}/n$ has expectation equal to 0 and its squared expectation converges to zero. Therefore, by Markov's inequality, it is $o_p(1)$. Furthermore, by stationarity,

$$E\sum_{i=1}^n \left(\sum_{j=0}^\infty \rho^j U_{i-1-j}\right)^2/n = E\left(\sum_{j=0}^\infty \rho^j U_{-j}\right)^2 = (1-\rho^2)^{-1}\sigma_U^2 \tag{A.15}$$

and (by straightforward but tedious calculations)

$$E\left(n^{-1}\sum_{i=1}^n \left(\sum_{j=0}^\infty \rho^j U_{i-1-j}\right)^2\right)^2 - (1-\rho^2)^{-2}\sigma_U^4 \to 0. \tag{A.16}$$

This establishes that the expression in the numerator of the third line of (A.14) converges in probability to zero, which shows that $v_n \to_p 1$. The other terms ξ_n and $(1-\rho^2)^{-1/2}\widehat{\sigma}_n$ are dealt with similarly but for brevity we do not provide the details. □

Proof of Lemma A.2. Note that $\rho = 1 - h_n/n$ and $h_n = O(1)$ imply that $\rho \to 1$. Recall h_n^* is defined by $\rho = \exp(-h_n^*/n)$. Hence, $\exp(-h_n^*/n) = \rho \to 1$ and $h_n^* = o(n)$. By a mean-value expansion of $\exp(-h_n^*/n)$ about 0,

$$0 = \rho - \rho = \exp(-h_n^*/n) - (1 - h_n/n) = h_n/n - \exp(-h_n^{**}/n)h_n^*/n, \tag{A.17}$$

for a h_n^{**} that satisfies $h_n^{**} = o(n)$ given that $h_n^* = o(n)$. Hence, $h_n - (1+o(1))h_n^* = 0$,

$$h_n^*/h_n \to 1, \tag{A.18}$$

and it suffices to prove the result with h_n^* in place of h_n.

Let $\{m_n : n \geq 1\}$ be a sequence such that $m_n h_n^*/n \to \infty$. By Assumption STAT (which holds because $\rho < 1$), we can write $(2h_n^*/n)^{1/2} Y_0^*/\sigma_U = A_{1n} + A_{2n}$ for $A_{1n} = (2h_n^*/n)^{1/2}\sum_{j=0}^{m_n} \rho^j U_{-j}/\sigma_U$ and $A_{2n} = (2h_n^*/n)^{1/2}\sum_{j=m_n+1}^{\infty} \rho^j U_{-j}/\sigma_U$. Note that $EA_{2n} = 0$ and

$$\begin{aligned} var(A_{2n}) &= (2h_n^*/n)\sum_{j=m_n+1}^\infty \rho^{2j} = (2h_n^*/n)\rho^{2(m_n+1)}/(1-\rho^2) \\ &= (2h_n^*/n)\rho^{2(m_n+1)}/((2h_n^*/n)(1+o(1))) = O(\exp(-2(m_n+1)h_n^*/n)) = o(1), \end{aligned} \tag{A.19}$$

where the third equality holds because $\rho^2 = \exp(-2h_n^*/n) = 1 - (2h_n^*/n)(1+o(1))$ by a mean value expansion and the last equality holds because $m_n h_n^*/n \to \infty$ by assumption. Therefore, $A_{2n} \to_p 0$. Next, write

$$A_{1n} = (2h_n^*/n)^{1/2}\sum_{i=-m_n}^0 \rho^{-i} U_i/\sigma_U = \sum_{i=-m_n}^0 X_{n,i} \tag{A.20}$$

with

$$X_{n,i} = (2h_n^*/n)^{1/2}\rho^{-i} U_i/\sigma_U. \tag{A.21}$$

To show $A_{1n} \to_d Z$ we apply the CLT in Corollary 3.1 in Hall and Heyde (1980) with $X_{n,i}$ and $\mathcal{F}_{n,i} = \{\varnothing, \Omega\}$, where Ω is the sample space. (Note that although this sum runs backwards, which differs from the result given in Hall and Heyde (1980), it can be converted easily into a sum that runs forward by changing variables via $j = i + m_n$. To keep the notation simpler, we do not do so.) To apply this CLT, we have to verify a Lindeberg condition and a variance condition.

We have $\sum_{i=-m_n}^0 \rho^{-2i} = (1-\rho^{2(m_n+1)})/(1-\rho^2)$, $\rho^{2(m_n+1)} = \exp(-2h_n^*(m_n+1)/n) \to 0$, and $n(1-\rho^2) = n(1-\rho)(1+\rho) = h_n(1+\rho) \to 2h$, which imply

$$2h_n^*\sum_{i=-m_n}^0 \rho^{-2i}/n \to 1. \tag{A.22}$$

To prove the Lindeberg condition, let $\varepsilon > 0$. We have

$$\begin{aligned}
&\sum_{i=-m_n}^{0} EX_{n,i}^2 I(|X_{n,i}|>\varepsilon) \\
&\leq (2h_n^*/n) \sum_{i=-m_n}^{0} \rho^{-2i} E((U_i^2/\sigma_U^2) I(2h_n^* U_i^2/\sigma_U^2 > n\varepsilon^2)) \\
&= (2h_n^*/n) \left[\sum_{i=0}^{m_n} \rho^{2i}\right] E((U_0^2/\sigma_U^2) I(2h_n^* U_0^2/\sigma_U^2 > n\varepsilon^2)) \\
&= O(1)o(1), \qquad \text{(A.23)}
\end{aligned}$$

where the inequality uses $\rho^{2i} > 1$ for $i < 0$, and the first equality holds because the U_{-i} have identical distributions. For the last equality, write $W_n = U_0^2/\sigma_U^2$. For any $\nu > 0$, $W_n I((2h_n^* W_n/(n\varepsilon^2))^\nu > 1) \leq W_n^{1+\nu}(2h_n^*/(n\varepsilon^2))^\nu$ and because of the restrictions in (2.12) we have $(2h_n^*/(n\varepsilon^2))^\nu EW_n^{1+\nu} = o(1)$ for some small $\upsilon > 0$. This and (A.22) establish the last equality in (A.23) and proves the Lindeberg condition.

The variance condition $\sum_{i=-m_n}^{0} E(X_{n,i}^2) - 1 = 0 = o(1)$ follows directly from (A.22). □

Proof of Lemma A.3. Parts of the proof reproduce the proof of Lemma 1 in Phillips (1987) and are included for completeness. To prove part (a), by (A.1), we have

$$\begin{aligned}
n^{-1/2} Y_{[nr]}^*/\sigma_U &= n^{-1/2}\widetilde{Y}_{[nr]}/\sigma_U + \exp(-h_n^*[nr]/n)(2h_n)^{-1/2}(2h_n/n)^{1/2} Y_0^*/\sigma_U \\
&\Rightarrow \int_0^r \exp(-(r-s)h)dW(s) + (2h)^{-1/2}\exp(-hr)Z = I_h^*(r), \qquad \text{(A.24)}
\end{aligned}$$

where the convergence result for the second summand follows from Lemma A.2. With regards to the first summand, note that for $X_n(r) \equiv n^{-1/2}\sum_{j=1}^{[nr]} U_j/\sigma_U$ we obtain

$$\begin{aligned}
n^{-1/2}\widetilde{Y}_{[nr]}/\sigma_U &= \sum_{j=1}^{[nr]} \exp(-h_n^*([nr]-j)/n) \int_{(j-1)/n}^{j/n} dX_n(s) \\
&= \sum_{j=1}^{[nr]} \int_{(j-1)/n}^{j/n} \exp(-h(r-s))dX_n(s) + o_p(1) \\
&= \int_0^r \exp(-h(r-s))dX_n(s) + o_p(1) \\
&= X_n(r) - h\int_0^r \exp(-h(r-s))X_n(s)ds + o_p(1) \\
&\Rightarrow W(r) - h\int_0^r \exp(-h(r-s))W(s)ds \\
&= \int_0^r \exp(-(r-s)h)dW(s), \qquad \text{(A.25)}
\end{aligned}$$

where the fourth equality holds by integration by parts, the second to last line holds by the functional CLT $X_n(\cdot) \Rightarrow W(\cdot)$ and the CMT, and the last line holds again by integration by parts. Parts (b)–(d) follow from triangular array WLLNs and a CLT using the uniform moment bounds imposed in (2.12). Parts (e) and (g) then follow immediately. For example, we have

$$n^{-3/2}\sum_{i=1}^{n} Y_{i-1}^*/\sigma_U = \int_0^1 n^{-1/2} Y_{[ns]}^*/\sigma_U ds \to_d \int I_h^*, \qquad \text{(A.26)}$$

where the convergence holds by the CMT and part (a).

To prove part (f), we use (A.1) to write

$$n^{-1}\sum_{i=1}^{n} Y_{i-1}^* U_i/\sigma_U^2 = n^{-1}\sum_{i=1}^{n}\widetilde{Y}_{i-1}U_i/\sigma_U^2 + (Y_0^*/\sigma_U)n^{-1}\sum_{i=1}^{n}\exp(-h_n^* i/n)U_i/\sigma_U, \qquad \text{(A.27)}$$

where $\widetilde{Y}_i \equiv \sum_{j=1}^{i} \exp(-h_n^*(i-j)/n)U_j = \exp(-h_n^*/n)\widetilde{Y}_{i-1} + U_i$ and $\widetilde{Y}_0 \equiv 0$. First, we consider the first summand on the rhs of (A.27). Let ρ denote $\exp(-h_n^*/n)$. We have

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\widetilde{Y}_i^2 &= n^{-1}\sum_{i=1}^{n}(\rho^2\widetilde{Y}_{i-1}^2 + 2\rho\widetilde{Y}_{i-1}U_i + U_i^2) \text{ or} \\ n^{-1}\widetilde{Y}_n^2 &= n^{-1}\sum_{i=1}^{n}((\rho^2-1)\widetilde{Y}_{i-1}^2 + 2\rho\widetilde{Y}_{i-1}U_i + U_i^2) \text{ or} \\ 2\rho n^{-1}\sum_{i=1}^{n}\widetilde{Y}_{i-1}U_i/\sigma_U^2 &= n^{-1}\widetilde{Y}_n^2/\sigma_U^2 + n(1-\rho^2)\left(n^{-2}\sum_{i=1}^{n}\widetilde{Y}_{i-1}^2/\sigma_U^2\right) - n^{-1}\sum_{i=1}^{n}U_i^2/\sigma_U^2. \end{aligned} \tag{A.28}$$

By (A.24), we have $n^{-1/2}\widetilde{Y}_{[nr]}/\sigma_U \Rightarrow \int_0^r \exp(-(r-s)h)dW(s) = I_h(r)$. This implies $n^{-1}\widetilde{Y}_n^2/\sigma_U^2 \to_d I_h^2(1)$ and $n^{-2}\sum_{i=1}^{n}\widetilde{Y}_{i-1}^2/\sigma_U^2 \to_d \int I_h^2$ using standard arguments. By part (c), $n^{-1}\sum_{i=1}^{n}U_i^2/\sigma_U^2 \to_p 1$. In addition, $\rho \to 1$ and $n(1-\rho^2) = -2h(1+o(1))$ (by a mean value expansion as in (A.6)). Using (A.28) and the preceding results, we obtain

$$n^{-1}\sum_{i=1}^{n}\widetilde{Y}_{i-1}U_i/\sigma_U^2 \to_d (I_h^2(1) - 2h\int I_h^2 - 1)/2 = \int I_h(r)dW(r), \tag{A.29}$$

where the last equality is taken from equation (8) in Phillips (1987) (which relies on the initial condition $\widetilde{Y}_0 = 0$).

Next, we consider the second summand in (A.27). By Lemma A.2, $(2h_n/n)^{1/2}Y_0^*/\sigma_U \to_d Z$. Below, we show that

$$n^{-1/2}\sum_{i=1}^{n}\exp(-h_n^* i/n)U_i/\sigma_U \to_d \int \exp(-hr)dW(r). \tag{A.30}$$

These results combine to show that the second summand in (A.27) converges in distribution to $(2h)^{-1/2}Z\int\exp(-hr)dW(r)$ (jointly with the convergence of the first summand). Hence, the lhs of (A.27) converges in distribution to

$$\int I_h(r)dW(r) + (2h)^{-1/2}Z\int\exp(-hr)dW(r) = \int I_h^*(r)dW(r), \tag{A.31}$$

where the equality uses $I_h^*(r) \equiv I_h(r) + (2h)^{-1/2}\exp(-hr)Z$ for $h > 0$. This gives the result of part (f).

For part (f), it remains to prove (A.30). An argument similar to that in (A.25) gives

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}\exp(-h_n^* i/n)U_i/\sigma_U &= \sum_{i=1}^{n}\exp(-h_n^* i/n)\int_{(i-1)/n}^{i/n} dX_n(r) \\ &= \sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\exp(-hr)dX_n(r) + o_p(1) \\ &= \int_0^1 \exp(-hr)dX_n(r) + o_p(1) \\ &= \exp(-h)X_n(1) + h\int_0^1\exp(-hr)X_n(r)dr + o_p(1) \\ &\Rightarrow \exp(-h)W(1) + h\int_0^1\exp(-hr)W(r)dr \\ &= \int_0^1\exp(-hr)dW(r), \end{aligned} \tag{A.32}$$

where the fourth equality holds by integration by parts, the second to last line holds by the functional CLT $X_n(\cdot) \Rightarrow W(\cdot)$ and the CMT, and the last line holds again by integration by parts.

To prove part (h) note that $\widetilde{Y}_{i-1}$ corresponds to an AR(1) model with zero initial condition. The same proof as above applies. □

Proof of Lemma A.4. To prove (a), note that because $Y_i^* = \sum_{j=0}^{\infty}\rho^j U_{i-j}$ and because U_{i-j} are i.i.d. with mean zero and variance σ_U^2, we have

$$EY_0^{*2} = \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\rho^{u+v}EU_{-u}U_{-v} = \sum_{u=0}^{\infty}\rho^{2u}\sigma_U^2 = (1-\rho^2)^{-1}\sigma_U^2. \tag{A.33}$$

To prove (b), note first that because $\rho \to 1$ it follows that $n^{-1}(1-\rho)^{1/2}X_1'X_2 = n^{-1}(1-\rho)^{1/2}\sum_{i=1}^n Y_{i-1}^* + o_p(1)$. By Markov's inequality it is enough to show that $n^{-2}(1-\rho)E(\sum_{i=1}^n Y_{i-1}^*)^2 = o(1)$. Writing $Y_i^* = \sum_{j=0}^\infty \rho^j U_{i-j}$, we have

$$E(\sum_{i=1}^n Y_{i-1}^*)^2 = \sum_{i,k=1}^n \sum_{j,l=0}^\infty \rho^{j+l} EU_{i-1-j}U_{k-1-l} \le 2\sigma_U^2 \sum_{i\ge k=1}^n \sum_{j=i-k}^\infty \rho^{k-i+2j} = 2\sigma_U^2 \sum_{i\ge k=1}^n \rho^{i-k} \sum_{l=0}^\infty \rho^{2l}, \tag{A.34}$$

where the inequality holds by the U_i being mean zero and i.i.d. and because the rhs sum is over $k = 1,\ldots,n$ and $i = k,\ldots,n$, rather than $i, k = 1,\ldots,n$, and the second equality uses the change of coordinates $l = j-(i-k)$. Because $\sum_{l=0}^\infty \rho^{2l} = O((1-\rho)^{-1})$ and because $\sum_{i\ge k=1}^n \rho^{i-k} = O(n(1-\rho)^{-1})$ (which follows from another change of variables $l = i-k$) the claim $n^{-2}(1-\rho)E(\sum_{i=1}^n Y_{i-1}^*)^2 = o(1)$ follows from $n(1-\rho)\to\infty$.

Next, we prove part (c) of the Lemma. By parts (a)–(b) and Markov's inequality, it is enough to show that

$$E\left(\frac{n^{-1}\sum_{i=1}^n Y_{i-1}^{*2} - (1-\rho^2)^{-1}\sigma_U^2}{(1-\rho^2)^{-1}\sigma_U^2}\right)^2 \to 0. \tag{A.35}$$

By part (a), the left hand side in (A.35) equals

$$\begin{aligned}
&(1-\rho^2)^2\sigma_U^{-4}\left(n^{-2}\sum_{i,j=1}^n EY_{i-1}^{*2}Y_{j-1}^{*2} - (1-\rho^2)^{-2}\sigma_U^4\right)\\
&= (1-\rho^2)^2\sigma_U^{-4}\left(n^{-2}\sum_{i,j=1}^n E\left[\sum_{l=0}^\infty \rho^l U_{i-1-l}\right]^2\left[\sum_{k=0}^\infty \rho^k U_{j-1-k}\right]^2 - (1-\rho^2)^{-2}\sigma_U^4\right) \qquad \text{(A.36)}\\
&= (1-\rho^2)^2\sigma_U^{-4}\times\\
&\quad\left(n^{-2}\sum_{i,j=1}^n \sum_{l_1,l_2=0}^\infty \sum_{k_1,k_2=0}^\infty \rho^{l_1+l_2+k_1+k_2}E(U_{i-1-l_1}U_{i-1-l_2}U_{j-1-k_1}U_{j-1-k_2}) - (1-\rho^2)^{-2}\sigma_U^4\right),
\end{aligned}$$

where the first equality uses $Y_i^* = \sum_{j=0}^\infty \rho^j U_{i-j}$. Any expectation in the third line is zero unless (i) all the indices on the four innovation terms coincide or (ii) there are two groups of two indices that each coincide. In case (i) we must have $l_1 = l_2$, $k_1 = k_2$, and $i-1-l_1 = j-1-k_1$. The sum of all the expectations (assuming fourth moments of the innovations are uniformly bounded) leads to an expression similar to the one in (A.34) that is of order $o(1)$ once normalized by $(1-\rho^2)^2 n^{-2}$. In case (ii) we must have (ii1) $(i-1-l_1 = i-1-l_2$ and $j-1-k_1 = j-1-k_2)$ or (ii2) $(i-1-l_1 = j-1-k_1$ and $i-1-l_2 = j-1-k_2)$ or (ii3) $(i-1-l_1 = j-1-k_2$ and $i-1-l_2 = j-1-k_1)$. Consider (ii1) first.

$$\begin{aligned}
&n^{-2}\sum_{i,j=1}^n \sum_{l_1,l_2=0}^\infty \sum_{\substack{k_1,k_2=0\\ case\ ii1)}}^\infty \rho^{l_1+l_2+k_1+k_2}E(U_{i-1-l_1}U_{i-1-l_2}U_{j-1-k_1}U_{j-1-k_2})\\
&= \sigma_U^4 n^{-2}\sum_{i,j=1}^n \sum_{\substack{k,l=0\\ i-l\ne j-k}}^\infty \rho^{2l+2k}\\
&= (1-\rho^2)^{-2}\sigma_U^4 - \sigma_U^4 n^{-2}\sum_{i,j=1}^n \sum_{\substack{k,l=0\\ i-l=j-k}}^\infty \rho^{2l+2k}\\
&= (1-\rho^2)^{-2}\sigma_U^4 - o((1-\rho^2)^{-2}), \qquad \text{(A.37)}
\end{aligned}$$

where the last equality follows again by proceeding as in (A.34). Therefore, to finish the proof of part (c) it is enough to show that the sum of all expectations in cases (ii2) and (ii3) are of order $o((1-\rho^2)^{-2}n^2)$. This is easily done along the same steps as above.

Finally, we now prove part (d). We have

$$(1-\rho^2)^{1/2}n^{-1}U'X_1 = (1-\rho^2)^{1/2}n^{-1}\sum_{i=1}^n U_iY_{i-1}^* + o_p(1). \tag{A.38}$$

Therefore, by Markov's inequality it is enough to establish that

$$(1-\rho^2)n^{-2}E(\sum_{i=1}^n U_iY_{i-1}^*)^2 = (1-\rho^2)n^{-2}\sum_{i,j=1}^n EU_iU_jY_{i-1}^*Y_{j-1}^* \to 0. \tag{A.39}$$

We have

$$(1-\rho^2)n^{-2}\sum_{i,j=1}^n EU_iU_jY_{i-1}^*Y_{j-1}^* = O((1-\rho^2)n^{-1})EY_0^{*2} \tag{A.40}$$

where the equality holds because the expectation on the left hand side is equal to zero whenever $i \neq j$, $EU_i^2 Y_{i-1}^{*2} = \sigma_U^2 EY_{i-1}^{*2}$, and because of stationarity. The result then follows from part (a) of the lemma. □

Proof of Lemma A.5. (a) By Markov's inequality, is enough to show that $\sum_{i=1}^n E(\zeta_i^2 1(|\zeta_i| > \delta)) \to 0$ for any $\delta > 0$. We have

$$\sum_{i=1}^n E(\zeta_i^2 1(|\zeta_i| > \delta)) = nE(\zeta_i^2 1(|\zeta_i| > \delta)) \leq n\delta^{-2} E(\zeta_1^4) = O(n^{-1}(1-\rho)^2) EY_0^{*4}, \tag{A.41}$$

where the last equality uses the moment restrictions imposed in (2.12). Furthermore, $EY_0^{*4} = \sum_{u,v,s,t=0}^{\infty} \rho^{u+v+s+t} EU_{-u}U_{-v}U_{-s}U_{-t}$. The contributions of all summands for which at least two of the indices u, v, s, t are the same is $o(n(1-\rho)^{-2})$. For example, suppose $u = v$, then $\sum_{u,s,t=0}^{\infty} \rho^{2u+s+t} = O((1-\rho)^{-3})$ which is indeed $o(n(1-\rho)^{-2})$ because $n(1-\rho) \to \infty$.

(b) By straightforward calculations, we have

$$E\left(\sum_{i=1}^n E(\zeta_i^2 | \mathcal{G}_{i-1})\right)^2 = \frac{n^{-2}}{(1-\rho^2)^{-2}\sigma_U^4} \sum_{i,j=1}^n EY_{i-1}^{*2} Y_{j-1}^{*2} \tag{A.42}$$

and thus

$$E\left(\sum_{i=1}^n E(\zeta_i^2 | \mathcal{G}_{i-1}) - 1\right)^2 = \frac{\sum_{i,j=1}^n EY_{i-1}^{*2} Y_{j-1}^{*2} - n^2(1-\rho^2)^{-2}\sigma_U^4}{n^2(1-\rho^2)^{-2}\sigma_U^4} \to 0, \tag{A.43}$$

where the convergence holds by (A.35). □

Proof of Lemma A.6. The proof is quite similar to the proof of Theorem 1 in Andrews and Guggenberger (2014, Appendix), but it corrects the proof given there that $c_{h_n}(\beta) \to c_h(\beta)$ as $h_n \to h \in [0, \infty]$. (That proof defines $L_n(x)$ to be the distribution function of $T_n(\rho_n)$, but $L_n(x)$ needs to be the distribution function $J_{h_n}(x)$ and it needs to be shown that $J_{h_n}(x) \to J_h(x)$ as $h_n \to h$ for all $x \in R$.) By Theorem 2.2 and Lemma 2.1, it is enough to verify Assumptions B1 and B2 with $CP^-(h) = CP^+(h) = 1 - \alpha$. To verify Assumption B1, recall that $\lambda_n = (\rho_n, F_n)$, $h_n(\lambda_n) = n(1-\rho_n)$, $H = [0, \infty]$, and the parameter space Λ is defined in (2.12). Consider a sequence $\{\lambda_n = (\rho_n, F_n) \in \Lambda : n \geq 1\}$ for which $h_n = h_n(\lambda_n) \to h \in H$, i.e., $\rho_n = 1 - h_n/n$ and $h_n \to h \in [0, \infty]$. Recall the definition of CI_n in (2.17) . We have $CP_n(\lambda_n) = P_{\lambda_n}(\rho_n \in CI_n) = P_{\lambda_n}(c_{h_n}(\alpha/2) \leq T_n(\rho_n) \leq c_{h_n}(1-\alpha/2))$. By (2.15), we have $T_n(\rho_n) \to_d J_h$ under $\{\lambda_n \in \Lambda : n \geq 1\}$. In addition, $c_{h_n}(\beta) \to c_h(\beta)$ for $\beta = \alpha/2$ and $1-\alpha/2$. The latter is proved as follows using Lemma A.7: Because J_h is strictly increasing at its β-quantile, for any $\varepsilon > 0$, $J_{h_n}(c_h(\beta) - \varepsilon) \to J_h(c_h(\beta) - \varepsilon) < \beta$ and $J_{h_n}(c_h(\beta) + \varepsilon) \to J_h(c_h(\beta) + \varepsilon) > \beta$. This and the definition $c_{h_n}(\beta) = \inf\{x \in R : J_{h_n}(x) \geq \beta\}$ yield $1\{c_h(\beta) - \varepsilon \leq c_{h_n}(\beta) \leq c_h(\beta) + \varepsilon\} \to 1$ as $n \to \infty$ for any $\varepsilon > 0$, which implies that $c_{h_n}(\beta) \to c_h(\beta)$. Now, using $c_{h_n}(\beta) \to c_h(\beta)$, the definition of convergence in distribution, and the continuity of J_h, it follows that $CP_n(\lambda_n) \to 1 - \alpha$.

For Assumption B2, assume we are given $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ for a subsequence $\{p_n\}$ of $\{n\}$ such that $h_{p_n}(\lambda_{p_n}) \to h \in H$. Define $\{\lambda_n^* : n \geq 1\}$ by (i) $\lambda_{p_n}^* = \lambda_{p_n}$ $\forall n \geq 1$, (ii) when $h < \infty$ and $m \neq p_n$, define $\lambda_m^* = (1 - h/m, F^*)$, and (iii) when $h = \infty$ and $m \neq p_n$, define $\lambda_m^* = (0, F^*)$, where F^* is the distribution such that $\{U_i : i = 0, \pm 1, \pm 2, \ldots\}$ are i.i.d., standard normal. Then, $\lambda_n^* \in \Lambda$ for all $n \geq 1$ and by construction $h_n(\lambda_n^*) \to h \in H$. This verifies Assumption B2 and completes the proof. □

Proof of Lemma A.7. The only two nontrivial cases are $h = 0$ and $h = \infty$.

Case $h = \infty$. First, we consider the case $h = \infty$. We need to show

$$\int_0^1 I_{D,h_n}^*(r) dW(r) \Big/ \left(\int_0^1 I_{D,h_n}^*(r)^2 dr\right)^{1/2} \to_d N(0,1) \tag{A.44}$$

for any sequence $h_n \to \infty$. Standardizing the numerator of (A.44) by $\sqrt{2h_n}$, we show the following convergence result:

$$\begin{aligned} &\sqrt{2h_n} \int_0^1 I_{D,h_n}^*(r) dW(r) \\ &= \sqrt{2h_n} \int_0^1 I_{h_n}(r) dW(r) + Z \int_0^1 \exp(-h_n r) dW(r) \\ &\quad - \sqrt{2h_n} W(1) \int_0^1 I_{h_n}(s) ds - ZW(1) \int_0^1 \exp(-h_n s) ds \\ &\to_d N(0,1) \text{ as } h_n \to \infty. \end{aligned} \tag{A.45}$$

The first summand on the second line of (A.45) converges in distribution to $N(0,1)$ by Lemma 2(b) in Phillips (1987).[23] To complete the verification of the convergence in (A.45), we show that the other summands on the second and third lines of (A.45) converge in probability to zero. By integration by parts, we have

$$\int_0^1 \exp(-h_n r)dW(r) = \exp(-h_n)W(1) + \int_0^1 h_n \exp(-h_n r)W(r)dr. \tag{A.46}$$

We now show that

$$\int_0^1 h_n \exp(-h_n r)W(r)dr \to_p 0 \tag{A.47}$$

as $h_n \to \infty$. Write

$$\int_0^1 h_n \exp(-h_n r)W(r)dr = \int_0^{h_n^{-1/2}} h_n \exp(-h_n r)W(r)dr + \int_{h_n^{-1/2}}^1 h_n \exp(-h_n r)W(r)dr. \tag{A.48}$$

The absolute value of the first summand is bounded by $\int_0^{h_n^{-1/2}} h_n \exp(-h_n r)dr \sup_{r\in[0,h_n^{-1/2}]}|W(r)|$. We have

$$\int_0^{h_n^{-1/2}} h_n \exp(-h_n s)ds = h_n(-\exp(-h_n s)/h_n)|_0^{h_n^{-1/2}} = -\exp(-h_n^{1/2}) + 1 \to 1 \tag{A.49}$$

as $h_n \to \infty$. Furthermore, $\sup_{r\in[0,h_n^{-1/2}]} W(r) = o(1)$ a.s. as $h_n \to \infty$ because Brownian motion has continuous sample paths a.s. and $W(0) = 0$ a.s. Hence, the first summand on the rhs of (A.48) is $o_p(1)$. Regarding the second summand on the rhs of (A.48), using the Cauchy Schwarz inequality, we obtain

$$\left(\int_{h_n^{-1/2}}^1 h_n \exp(-h_n r)W(r)dr\right)^2 \le \int_{h_n^{-1/2}}^1 h_n^2 \exp(-2h_n r)dr \int_{h_n^{-1/2}}^1 W(r)^2 dr. \tag{A.50}$$

We have

$$h_n^2 \int_{h_n^{-1/2}}^1 \exp(-2h_n r)dr = h_n^2(-.5h_n^{-1}\exp(-2h_n r))|_{h_n^{-1/2}}^1 = -.5h_n \exp(-2h_n) + .5h_n \exp(-2h_n^{1/2}), \tag{A.51}$$

which converges to zero as $h_n \to \infty$. Hence, the second summand on the rhs of (A.48) is $o_p(1)$, which implies that second summand on the second line of (A.45) is $o_p(1)$.

Regarding the first summand on the third line of (A.45), from line 3 on p. 540 of Phillips (1987), we have

$$\sqrt{2h_n}\int_0^1 I_{h_n}(s)ds \sim N(0, v_n), \text{ where}$$
$$v_n \equiv 2h_n^{-1} - h_n^{-2}(\exp(-2h_n) - 4\exp(-h_n) + 3) \to 0 \text{ as } h_n \to \infty. \tag{A.52}$$

Therefore, the first summand on the third line of (A.45) is $o_p(1)$ as $h_n \to \infty$.

For the second summand on the third line of (A.45), we have

$$\int_0^1 \exp(-h_n s)ds = (-\exp(-h_n s)/h_n)|_0^1 = -\exp(-h_n)/h_n + 1/h_n \to 0 \tag{A.53}$$

as $h_n \to \infty$. This completes the verification of the convergence result in (A.45).

Next, standardizing the denominator of (A.44) by $\sqrt{2h_n}$, we obtain

$$\sqrt{2h_n}\left(\int_0^1 I_{D,h_n}^*(r)^2 dr\right)^{1/2} = \left(2h_n \int_0^1 I_{h_n}^*(r)^2 dr - 2h_n\left(\int_0^1 I_{h_n}^*(s)ds\right)^2\right)^{1/2}. \tag{A.54}$$

The first summand inside the square root on the rhs of (A.54) equals

$$2h_n \int_0^1 I_{h_n}(r)^2 dr + 2(2h_n)^{1/2} Z \int_0^1 I_{h_n}(r)\exp(-h_n r)dr + Z^2 \int_0^1 \exp(-2h_n r)dr. \tag{A.55}$$

By Lemma 2(a) in Phillips (1987), the first summand in (A.55) converges in probability to 1 as $h_n \to \infty$. By (A.53) with $2h_n$ in place of h_n, the third summand in (A.55) is $o_p(1)$ as $h_n \to \infty$. Using the results for the first and third summands in (A.55) and the Cauchy Schwarz inequality, the second summand in (A.55) is $o_p(1)$ as $h_n \to \infty$. Hence, the first summand inside the square root on the rhs of (A.54) converges in probability to 1.

[23] Note that the process $J_c(r)$ in Phillips (1987, p. 538) corresponds to $I_{-c}(r)$ in the current paper.

The second summand inside the square root on the rhs of (A.54) converges to zero in probability. To show this, note that the second summand equals

$$2h_n\left(\int_0^1 I_{h_n}^*(s)ds\right)^2 = \left((2h_n)^{1/2}\int_0^1 I_{h_n}(s)ds + Z\int_0^1 \exp(-h_n s)ds\right)^2. \quad \text{(A.56)}$$

The first summand inside the square on the rhs is $o_p(1)$ by the same argument as for the first summand on the third line of (A.45). The second summand inside the square on the rhs of (A.56) is $o_p(1)$ as $h_n \to \infty$ by (A.53). Hence, the standardized denominator in (A.54) converges in probability to one as $h_n \to \infty$. This concludes the verification of $J_{h_n} \to_d J_h$ for the case $h = \infty$.

Case $h = 0$. Next, we verify $J_{h_n} \to_d J_h$ when $h = 0$. Using integration by parts, we have

$$I_{h_n}(r) = \int_0^r \exp(-(r-s)h_n)dW(s) = W(r) - \int_0^r h_n \exp(-(r-s)h_n)W(s)ds. \quad \text{(A.57)}$$

Therefore, for the numerator in (A.44), we show the following convergence result:

$$\begin{aligned}
&\int_0^1 I_{D,h_n}^*(r)dW(r)\\
&= \int_0^1\left[W(r) - \int_0^r h_n\exp(-(r-s)h_n)W(s)ds + \frac{1}{\sqrt{2h_n}}\exp(-h_n r)Z\right]dW(r)\\
&\quad - \int_0^1\int_0^1\left[W(s) - \int_0^s h_n\exp(-(s-t)h_n)W(t)dt + \frac{1}{\sqrt{2h_n}}\exp(-h_n s)Z\right]dsdW(r)\\
&\to_d \int_0^1 W(r)dW(r) - \int_0^1\int_0^1 W(s)dsdW(r) = \int_0^1 I_{D,0}^*(r)dW(r) \text{ as } h_n \to 0. \quad \text{(A.58)}
\end{aligned}$$

By a mean-value expansion around $h_n = 0$, $\sup_{r,s\in[0,1]}|\exp(-(r-s)h_n) - 1| \to 0$ as $h_n \to 0$. In consequence, the middle summands in lines two and three of (A.58) are $o_p(1)$. Furthermore,

$$\begin{aligned}
&\int_0^1 \frac{1}{\sqrt{2h_n}}\exp(-h_n r)ZdW(r) - \int_0^1\int_0^1 \frac{1}{\sqrt{2h_n}}\exp(-h_n s)ZdsdW(r)\\
&= Z\frac{1}{\sqrt{2h_n}}\int_0^1[\exp(-h_n r) - 1 + O(h_n)]dW(r)\\
&= Z\frac{1}{\sqrt{2h_n}}[\exp(-h_n)W(1) + h_n\int_0^1 \exp(-h_n r)W(r)dr - W(1) + O_p(h_n)]\\
&= Z\frac{1}{\sqrt{2h_n}}[h_n\int_0^1 \exp(-h_n r)W(r)dr + O_p(h_n)]\\
&= o_p(1) \text{ as } h_n \to 0, \quad \text{(A.59)}
\end{aligned}$$

where the $O(h_n)$ term does not depend on r, the first equality uses the two equalities in (A.53) and $-\exp(-h_n)/h_n + 1/h_n = -1 + O(h_n)$ by a two-term Taylor expansion of $\exp(-h_n)$ about 0, and the second equality uses (A.46). Hence, the convergence in (A.58) holds.

Next, we deal with the denominator $(\int_0^1 I_{D,h_n}^*(r)^2dr)^{1/2} = (\int_0^1(I_{h_n}^*(r) - \int_0^1 I_{h_n}^*(s)ds)^2dr)^{1/2}$ in (A.44) for the case $h = 0$. We need to show that it converges in distribution to $(\int_0^1(W(r) - \int_0^1 W(s)ds)^2dr)^{1/2}$ jointly with the numerator. We have

$$\begin{aligned}
&\int_0^1\left(I_{h_n}^*(r) - \int_0^1 I_{h_n}^*(s)ds\right)^2 dr\\
&= \int_0^1\left(I_{h_n}(r) - \int_0^1 I_{h_n}(s)ds + Z(2h_n)^{-1/2}\left[\exp(-h_n r) - \int_0^1\exp(-h_n s)ds\right]\right)^2 dr \quad \text{(A.60)}\\
&= \int_0^1\left(I_{h_n}(r) - \int_0^1 I_{h_n}(s)ds + Z(2h_n)^{-1/2}\left[\exp(-h_n r) + h_n^{-1}\exp(-h_n) - h_n^{-1}\right]\right)^2 dr,
\end{aligned}$$

where the last line uses (A.53). Using (A.57) and $\sup_{r,s\in[0,1]} |\exp(-(r-s)h_n) - 1| \to 0$ as $h_n \to 0$, it follows that $I_{h_n}(r) \Rightarrow W(r)$ as $h_n \to 0$. Therefore, by the CMT, we obtain

$$\int_0^1 (I_{h_n}(r) - \int_0^1 I_{h_n}(s)ds)^2 dr \to_d \int_0^1 (W(r) - \int_0^1 W(s)ds)^2 dr. \quad (A.61)$$

Thus, it suffices to show that

$$\int_0^1 (2h_n)^{-1} \left(\exp(-h_n r) + h_n^{-1}\exp(-h_n) - h_n^{-1}\right)^2 dr \to 0 \text{ as } h_n \to 0. \quad (A.62)$$

By first-order and second-order Taylor expansions, for some intermediate values $\xi_n \in [-h_n r, 0]$ and $\xi_n' \in [-h_n, 0]$, we have

$$\begin{aligned} & (2h_n)^{-1/2}(\exp(-h_n r) + h_n^{-1}(\exp(-h_n) - 1)) \\ = & (2h_n)^{-1/2} \left(1 - h_n r \exp(\xi_n) + h_n^{-1}(-h_n + h_n^2 \exp(\xi_n')/2)\right) \\ = & (2h_n)^{-1/2} \left(-h_n r \exp(\xi_n) + h_n \exp(\xi_n')/2\right) \\ = & O(h_n^{1/2}) \end{aligned} \quad (A.63)$$

uniformly in $r \in [0, 1]$. Therefore, the integrand in (A.62) goes to zero as $h_n \to 0$ uniformly in $r \in [0, 1]$ and (A.62) holds. This completes the verification of $J_{h_n} \to_d J_h$ for the case $h = 0$. □

A.2. Conditional likelihood ratio test with weak instruments

Proof of Lemma 3.1. We start by proving the result for the full sequence $\{n\}$, rather than a subsequence $\{p_n\}$. Then, we note that the same proof goes through with p_n in place of n.

Let $\{\lambda_n\}$ be a sequence in Λ such that $h_n(\lambda_n) \to h = (h_1, h_2, h_{31,\ldots}, h_{34}) \in H$. All results stated below are "under $\{\lambda_n\}$." Note that λ_n determines F_n (because $\lambda_{5F_n} = F_n$ by (3.13)) and we write expectations under λ_n as E_{F_n}. Define

$$\begin{aligned} h_1 &= \lim_{n\to\infty} n^{1/2}\|\pi_n\|, \quad h_2 = \lim_{n\to\infty} \pi_n/\|\pi_n\|, \quad h_{31} = \lim_{n\to\infty} E_{F_n} Z_i^* Z_i^{*\prime}, \\ h_{32} &= \lim_{n\to\infty} \Omega_{F_n} = \lim_{n\to\infty} E_{F_n} Z_i^* Z_i^{*\prime} u_i^2, \quad h_{33} = \lim_{n\to\infty} \Sigma_{F_n} = \lim_{n\to\infty} E_{F_n} Z_i^* Z_i^{*\prime} v_i^2, \text{ and} \\ h_{34} &= \lim_{n\to\infty} \Gamma_{F_n} = \lim_{n\to\infty} E_{F_n} Z_i^* Z_i^{*\prime} u_i v_i. \end{aligned} \quad (A.64)$$

Below we use the following result that holds by a WLLN for a triangular array of row-wise i.i.d. random variables using the moment conditions in Λ:

$$n^{-1}\sum_{i=1}^n A_{1i}A_{2i}A_{3i}A_{4i} - E_{F_n} A_{1i}A_{2i}A_{3i}A_{4i} \to_p 0, \quad (A.65)$$

where A_{1i} and A_{2i} consist of any elements of Z_i, X_i, or Z_i^* and A_{3i} and A_{4i} consist of any elements of Z_i, X_i, u_i, v_i, or 1.

Using the definitions in (3.6) and $y_{1i}^\perp = y_{2i}^\perp \theta_0 + u_i^\perp$, we obtain

$$\begin{aligned} & n^{-1}\sum_{i=1}^n g_i g_i' = n^{-1}\sum_{i=1}^n Z_i^\perp Z_i^{\perp\prime} u_i^{\perp 2} = n^{-1}\sum_{i=1}^n Z_i^* Z_i^{*\prime} u_i^2 + o_p(1), \text{ where} \\ & Z_i^\perp = Z_i - (n^{-1}\sum_{i=1}^n Z_i X_i')(n^{-1}\sum_{i=1}^n X_i X_i')^{-1} X_i, \\ & Z_i^* = Z_i - (E_{F_n} Z_i X_i')(E_{F_n} X_i X_i')^{-1} X_i, \text{ and} \\ & u_i^\perp = u_i - (n^{-1}\sum_{i=1}^n u_i X_i')(n^{-1}\sum_{i=1}^n X_i X_i')^{-1} X_i, \end{aligned} \quad (A.66)$$

where the second equality in the first line holds by (A.65) with $A_{1i}, \ldots, A_{4i}$ including elements of Z_i or X_i, Z_i or X_i, X_i, u_i, or 1, and X_i, u_i, or 1, respectively, $E_{F_n} X_i u_i = 0$, and some calculations, the second and fourth lines hold by (3.3), and the third line follows from (3.5) by noting that Z_i^* in (3.5) depends on E_F and in the present case F is F_n.

By Lyapunov's triangular CLT, we obtain

$$n^{1/2}\widehat{g} = n^{-1/2}\sum_{i=1}^{n} Z_i^{\perp}u_i^{\perp} = n^{-1/2}Z'M_X u = n^{-1/2}\sum_{i=1}^{n} Z_i^{\perp}u_i$$
$$= n^{-1/2}\sum_{i=1}^{n} Z_i^{*}u_i + o_p(1) \rightarrow_d N_h \sim N(0, h_{32}), \quad (A.67)$$

where the fourth equality holds using (A.65) and some calculations and the convergence uses $E_{F_n}\overline{Z}_i u_i = 0$, the moment conditions in Λ, and (A.64).

Eqs. (A.64)–(A.67) give

$$\widehat{\Omega} = n^{-1}\sum_{i=1}^{n} g_i g_i' - \widehat{g}\widehat{g}' = n^{-1}\sum_{i=1}^{n} Z_i^{*}Z_i^{*\prime}u_i^2 + o_p(1) \rightarrow_p h_{32}. \quad (A.68)$$

To obtain analogous results for $\widehat{\Sigma}$ and $\widehat{\Gamma}$, we write

$$\widehat{v} = M_{\overline{Z}}y_2 = M_{\overline{Z}}v, \;\; \widehat{v}_i = v_i - (n^{-1}\sum_{i=1}^{n} v_i\overline{Z}_i)(n^{-1}\sum_{i=1}^{n}\overline{Z}_i\overline{Z}_i')^{-1}\overline{Z}_i, \quad (A.69)$$
$$\widehat{u} = M_X(y_1 - y_2\theta_0) = M_X u, \text{ and } \widehat{u}_i = u_i - (n^{-1}\sum_{i=1}^{n} u_i X_i)(n^{-1}\sum_{i=1}^{n} X_i X_i')^{-1}X_i.$$

This gives

$$\widehat{\Sigma} = n^{-1}\sum_{i=1}^{n} Z_i^{\perp}Z_i^{\perp\prime}\widehat{v}_i^2 = n^{-1}\sum_{i=1}^{n} Z_i^{*}Z_i^{*\prime}v_i^2 + o_p(1) \rightarrow_p h_{33}, \text{ and}$$
$$\widehat{\Gamma} = n^{-1}\sum_{i=1}^{n} Z_i^{\perp}Z_i^{\perp\prime}\widehat{u}_i\widehat{v}_i = n^{-1}\sum_{i=1}^{n} Z_i^{*}Z_i^{*\prime}u_i v_i + o_p(1) \rightarrow_p h_{34}, \quad (A.70)$$

where the equalities in both lines use (A.65), (the second and third lines of) (A.66), (A.67), (A.69), and some calculations, and the convergence in both lines holds by (A.64) and (A.65).

Eqs. (A.68) and (A.70) and the condition $\lambda_{\min}(\Omega_F) \geq \delta > 0$ in Λ give

$$\widehat{\Psi} = \widehat{\Sigma} - \widehat{\Gamma}\widehat{\Omega}^{-1}\widehat{\Gamma} \rightarrow_p h_{33} - h_{34}h_{32}^{-1}h_{34} \equiv \Psi_h. \quad (A.71)$$

Case $\boldsymbol{h}_1 < \infty$. Using the results just established, we now prove the result of the Lemma for the case $h_1 < \infty$. We have

$$n^{-1}\sum_{i=1}^{n} G_i g_i' = n^{-1}\sum_{i=1}^{n} Z_i^{\perp}Z_i^{\perp\prime}(\pi_n' Z_i^{\perp} + v_i^{\perp})u_i^{\perp} = n^{-1}\sum_{i=1}^{n} Z_i^{*}Z_i^{*\prime}v_i u_i + o_p(1) \rightarrow_p h_{34}, \quad (A.72)$$

where the first equality uses $y_{2i}^{\perp} = Z_i^{\perp\prime}\pi_n + v_i^{\perp}$, the second equality holds using (i) $\limsup_{n\to\infty}\|\pi_n\| < \infty$, (ii) $E_{F_n}\overline{Z}_i(u_i, v_i) = 0$, (iii) (A.65) with A_{1i}, A_{2i}, A_{3i}, and A_{4i} including elements of Z_i or X_i, Z_i or X_i, X_i or u_i, and Z_i, X_i, or v_i, respectively, and (iv) some calculations, and the convergence uses (A.64) and (A.65).

By (A.64), we have

$$n^{1/2}E_{F_n}Z_i^{*}Z_i^{*\prime}\pi_n = n^{1/2}\|\pi_n\|E_{F_n}Z_i^{*}Z_i^{*\prime}(\pi_n/\|\pi_n\|) \rightarrow h_1 h_{31}h_2. \quad (A.73)$$

Using this, we obtain

$$n^{1/2}\widehat{G} = n^{-1/2}\sum_{i=1}^{n} Z_i^{\perp}y_{2i}^{\perp} = n^{-1/2}Z'M_X y_2$$
$$= n^{-1}Z'M_X Z(n^{1/2}\pi_n) + n^{-1/2}Z'M_X v$$
$$= n^{-1}\sum_{i=1}^{n} Z_i^{*}Z_i^{*\prime}(n^{1/2}\pi_n) + n^{-1/2}\sum_{i=1}^{n} Z_i^{*}v_i + o_p(1)$$
$$= h_1 h_{31}h_2 + n^{-1/2}\sum_{i=1}^{n} Z_i^{*}v_i + o_p(1), \quad (A.74)$$

where the fourth equality holds using $\limsup_{n\to\infty} n^{1/2}\|\pi_n\| < \infty$ (since $h_1 < \infty$), (A.65), and some calculations, and the last equality holds by (A.65) and (A.73) .

Using the definition of $\widehat{D}$ in (3.7), combined with (A.67), (A.68), (A.72) , and (A.74) yields

$$n^{1/2}\widehat{D} = n^{1/2}\widehat{G} - [n^{-1}\sum_{i=1}^{n} G_i g_i' - \widehat{G}\widehat{g}']\widehat{\Omega}^{-1}n^{1/2}\widehat{g}$$
$$= h_1 h_{31} h_2 + n^{-1/2}\sum_{i=1}^{n} Z_i^* v_i - h_{34}h_{32}^{-1}n^{-1/2}\sum_{i=1}^{n} Z_i^* u_i + o_p(1)$$
$$= h_1 h_{31} h_2 + [-h_{34}h_{32}^{-1} : I_{d_Z}]n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} Z_i^* u_i \\ Z_i^* v_i \end{pmatrix} + o_p(1), \quad \text{(A.75)}$$

where the second equality uses the condition $\lambda_{\min}(\Omega_F) \geq \delta > 0$ in Λ.

Combining (A.67) and (A.75) gives

$$\begin{pmatrix} n^{1/2}\widehat{g} \\ n^{1/2}\widehat{D} \end{pmatrix} = \begin{pmatrix} 0 \\ h_1 h_{31} h_2 \end{pmatrix} + \begin{bmatrix} I_{d_Z} & 0 \\ -h_{34}h_{32}^{-1} & I_{d_Z} \end{bmatrix} n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} Z_i^* u_i \\ Z_i^* v_i \end{pmatrix} + o_p(1)$$
$$\to_d \begin{pmatrix} N_h \\ D_h \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ h_1 h_{31} h_2 \end{pmatrix}, \begin{bmatrix} I_{d_Z} & 0 \\ -h_{34}h_{32}^{-1} & I_{d_Z} \end{bmatrix}\begin{bmatrix} h_{32} & h_{34} \\ h_{34} & h_{33} \end{bmatrix}\begin{bmatrix} I_{d_Z} & -h_{32}^{-1}h_{34} \\ 0 & I_{d_Z} \end{bmatrix}\right)$$
$$= N\left(\begin{pmatrix} 0 \\ h_1 h_{31} h_2 \end{pmatrix}, \begin{bmatrix} h_{32} & 0 \\ 0 & \Psi_h \end{bmatrix}\right), \text{ where } \Psi_h = h_{33} - h_{34}h_{32}^{-1}h_{34}. \quad \text{(A.76)}$$

The convergence in (A.76) holds by Lyapunov's triangular array CLT using the fact that $E_{F_n}\overline{Z}_i(u_i, v_i) = 0$ implies that $E_{F_n}Z_i^* u_i = E_{F_n}Z_i^* v_i = 0$, the moment conditions in Λ, and (A.64). In sum, (A.76) shows that $n^{1/2}\widehat{g}$ and $n^{1/2}\widehat{D}$ are asymptotically independent with asymptotic distributions $N_h \sim N(0, h_{32})$ and $D_h \sim N(h_1 h_{31} h_2, \Psi_h)$.

By the definition of Λ, $\lambda_{\min}(\Psi_h) \geq \delta > 0$. Hence, with probability one, $D_h \neq 0$. This, (A.76), and the CMT give

$$(\widehat{D}'\widehat{\Omega}^{-1}\widehat{D})^{-1/2}\widehat{D}'\widehat{\Omega}^{-1}n^{1/2}\widehat{g} \to_d (D_h' h_{32}^{-1} D_h)^{-1/2} D_h' h_{32}^{-1} N_h \equiv \zeta_1 \sim N(0, 1) \text{ and} \quad \text{(A.77)}$$
$$LM \to_d LM_h = N_h' h_{32}^{-1/2} P_{h_{32}^{-1/2}D_h} h_{32}^{-1/2} N_h = (D_h' h_{32}^{-1} D_h)^{-1}(D_h' h_{32}^{-1} N_h)^2 = \zeta_1^2 \sim \chi_1^2,$$

where $\zeta_1 \sim N(0, 1)$ because its conditional distribution given D_h is $N(0, 1)$ a.s.

We can write

$$AR = LM + J, \text{ where } J = n\widehat{g}'\widehat{\Omega}^{-1/2}M_{\widehat{\Omega}^{-1/2}\widehat{D}}\widehat{\Omega}^{-1/2}\widehat{g}. \quad \text{(A.78)}$$

Using (A.68), (A.71), (A.76), and the CMT, we have

$$J \to_d J_h = N_h' h_{32}^{-1/2} M_{h_{32}^{-1/2}D_h} h_{32}^{-1/2} N_h \sim \chi^2_{d_Z - 1} \text{ and}$$
$$W = n\widehat{D}'\widehat{\Psi}^{-1}\widehat{D} \to_d W_h = D_h' \Psi_h^{-1} D_h \sim \chi^2_{d_Z}. \quad \text{(A.79)}$$

Substituting (A.78) into the definition of CLR in (3.9) and using the convergence results in (A.77) and (A.79) (which hold jointly) and the CMT, we obtain

$$CLR \to_d CLR_h = \frac{1}{2}\left(LM_h + J_h - W_h + \sqrt{(LM_h + J_h - W_h)^2 + 4LM_h W_h}\right). \quad \text{(A.80)}$$

Next, we determine the asymptotic distribution of the CLR critical value. By definition, $c(1-\alpha, w)$ is the $(1-\alpha)$-quantile of the distribution of $clr(w)$ defined in (3.10). First, we show that $c(1-\alpha, w)$ is a continuous function of $w \in R_+$. To do so, consider a sequence $w_n \in R_+$ such that $w_n \to w \in R_+$. By the functional form of $clr(w)$, we have $clr(w_n) \to clr(w)$ as $w_n \to w$ a.s. Hence, by the bounded convergence theorem, for all continuity points y of $G_L(x) \equiv P(clr(w) \leq x)$, we have

$$L_n(y) \equiv P(clr(w_n) \leq y) = P(clr(w) + (clr(w_n) - clr(w)) \leq y)$$
$$\to P(clr(w) \leq y) = G_L(y). \quad \text{(A.81)}$$

The distribution function $G_L(x)$ is increasing at its $(1-\alpha)$ -quantile $c(1-\alpha, w)$. Therefore, by Andrews and Guggenberger (2010a, Lemma 5), it follows that $c(1-\alpha, w_n) \to_p c(1-\alpha, w)$. Because these quantities actually are nonrandom, we get $c(1-\alpha, w_n) \to c(1-\alpha, w)$. This establishes continuity.

From the continuity of $clr(w)$, (A.79), and (A.80), it follows that

$$CLR - c(1-\alpha, W) \to_d CLR_h - c(1-\alpha, W_h). \quad \text{(A.82)}$$

Therefore, by the definition of convergence in distribution, we have

$$P_{\theta_0,\lambda_n}(CLR > c(1-\alpha, W)) \to P(CLR_h > c(1-\alpha, W_h)), \quad \text{(A.83)}$$

where $P_{\theta_0,\lambda_n}(\cdot)$ denotes probability under λ_n when the true value of θ is θ_0. Now, conditional on $D_h = d$, W_h equals the constant $w \equiv d'\Psi_h^{-1}d$ and LM_h and J_h in (A.80) are independent (because they are quadratic forms in the normal vector

$h_{32}^{-1/2}N_h$ and $P_{h_{32}^{-1/2}D_h}M_{h_{32}^{-1/2}D_h}=0$) and are distributed as χ_1^2 and $\chi_{d_Z-1}^2$, respectively. Hence, the conditional distribution of CLR_h is the same as that of $clr(w)$, defined in (3.10), whose $(1-\alpha)$-quantile is $c(1-\alpha,w)$. This implies that the probability of the event $CLR_h > c(1-\alpha,W_h)$ in (A.83) conditional on $D_h=d$ equals α for all $d>0$. In consequence, the unconditional probability $P(CLR_h > c(1-\alpha,W_h))$ equals α as well. This completes the proof for the case $h_1<\infty$.

Case $h_1=\infty$: From here on, we consider the case where $h_1=\infty$. In this case, $\|\pi_n\|>0$ for all n large. Thus, we have

$$\|\pi_n\|^{-1}n^{-1}\sum_{i=1}^n G_ig_i = n^{-1}\sum_{i=1}^n Z_i^{\perp}Z_i^{\perp\prime}((\pi_n/\|\pi_n\|)'Z_i^{\perp}+v_i^{\perp}/\|\pi_n\|)u_i^{\perp}=O_p(1) \text{ and}$$

$$\|\pi_n\|^{-1}\widehat{G}=\|\pi_n\|^{-1}[n^{-1}Z'M_XZ\pi_n+n^{-1}Z'M_Xv]=n^{-1}\sum_{i=1}^n Z_i^*Z_i^{*\prime}(\pi_n/\|\pi_n\|)+o_p(1)$$

$$\to_p h_{31}h_2, \tag{A.84}$$

where the first equality in the first line uses the first equality in (A.72), the second equality in the first line uses (A.65), the moment conditions in Λ, $\|\pi_n/\|\pi_n\|\|=1$, and some calculations, the first equality in the second line uses the first two lines of (A.74), the second equality in the second line uses (A.65) and some calculations, and the convergence uses (A.64) and (A.65).

By the definition of $\widehat{D}$ and (A.84), we have

$$\|\pi_n\|^{-1}\widehat{D}=\|\pi_n\|^{-1}\widehat{G}-\|\pi_n\|^{-1}[n^{-1}\sum_{i=1}^n G_ig_i'-\widehat{G}\widehat{g}']\widehat{\Omega}^{-1}\widehat{g}\to_p h_{31}h_2, \tag{A.85}$$

where $\widehat{g}=o_p(1)$ by (A.67) and $\widehat{\Omega}^{-1}=O_p(1)$ by (A.68) and the condition $\lambda_{\min}(\Omega_F)\ge\delta>0$ in Λ.

Combining (A.67) and (A.85) gives

$$\begin{aligned}
&(\widehat{D}'\widehat{\Omega}^{-1}\widehat{D})^{-1/2}\widehat{D}'\widehat{\Omega}^{-1}n^{1/2}\widehat{g}=(\|\pi_n\|^{-1}\widehat{D}'\widehat{\Omega}^{-1}\|\pi_n\|^{-1}\widehat{D})^{-1/2}\|\pi_n\|^{-1}\widehat{D}'\widehat{\Omega}^{-1}n^{1/2}\widehat{g}\\
&\to_d (h_2'h_{31}h_{32}^{-1}h_{31}h_2)^{-1/2}h_2'h_{31}h_{32}^{-1}N_h\equiv\zeta_2\sim N(0,1) \text{ and}\\
&LM\to_d LM_h=N_h'h_{32}^{-1/2}P_{h_{32}^{-1/2}h_{31}h_2}h_{32}^{-1/2}N_h\\
&=(h_2'h_{31}h_{32}^{-1}h_{31}h_2)^{-1}(h_2'h_{31}h_{32}^{-1}N_h)^2=\zeta_2^2\sim\chi_1^2.
\end{aligned}\tag{A.86}$$

Analogously, $J\to_d N_h'h_{32}^{-1/2}M_{h_{32}^{-1/2}h_{31}h_2}h_{32}^{-1/2}N_h$ and, hence, $J=O_p(1)$.

From (A.85), $h_1=\lim_{n\to\infty}n^{1/2}\|\pi_n\|=\infty$, and $\|h_{31}h_2\|>0$, it follows that for all $K<\infty$,

$$P_{\theta_0,\lambda_n}(n^{1/2}\|\pi_n\|\cdot\|\pi_n\|^{-1}\|\widehat{D}\|>K)\to 1. \tag{A.87}$$

This, (A.71), and $\|\Psi_h\|<\infty$ (by the conditions in Λ) yield

$$P_{\theta_0,\lambda_n}(W>K)=P_{\theta_0,\lambda_n}(n\widehat{D}'\widehat{\Psi}^{-1}\widehat{D}>K)\to 1. \tag{A.88}$$

By (A.78) and some calculations, we have

$$(AR-W)^2+4LM\cdot W=(LM-J+W)^2+4LM\cdot J. \tag{A.89}$$

Substituting this into the expression for CLR in (3.9) gives

$$CLR=\frac{1}{2}\left(LM+J-W+\sqrt{(LM-J+W)^2+4LM\cdot J}\right). \tag{A.90}$$

Using a first-order expansion of the square-root expression in (A.90) about $(LM-J+W)^2$, we obtain

$$\sqrt{(LM-J+W)^2+4LM\cdot J}=LM-J+W+(1/2)\xi^{-1/2}4LM\cdot J \tag{A.91}$$

for an intermediate value ξ between $(LM-J+W)^2$ and $(LM-J+W)^2+4LM\cdot J$. By (A.86) and (A.88), $\xi^{-1/2}LM\cdot J\to_p 0$.

This, (A.86), (A.90), and (A.91) give

$$CLR=LM+o_p(1)\to_d\zeta_2^2\sim\chi_1^2. \tag{A.92}$$

Define $CLR_h(w)$ as CLR_h is defined in (A.80), but with w in place of W_h. We have $CLR_h(w)\to_d LM_h\sim\chi_1^2$ as $w\to\infty$ by the argument just given in (A.89)–(A.92). In consequence, the $(1-\alpha)$ -quantile $c(1-\alpha,w)$ satisfies $c(1-\alpha,w)\to\chi_{1,1-\alpha}^2$ as $w\to\infty$, where $\chi_{1,1-\alpha}^2$ is the $(1-\alpha)$-quantile of the χ_1^2 distribution. Combining this with (A.88) gives

$$c(1-\alpha,W)\to_p\chi_{1,1-\alpha}^2. \tag{A.93}$$

The result of Lemma 3.1 for the case $h_1=\infty$ follows from (A.92), (A.93), and the definition of convergence in distribution. □

A.3. Grid bootstrap CI in an AR(k) model

Proof of Lemma 4.1. We start by proving the result for the full sequence $\{n\}$ rather than the subsequence $\{p_n\}$. Then, we note that the same proof goes through with p_n in place of n. The t statistic for ρ_1 is invariant to μ_0 and μ_1. Hence, without loss of generality, we assume $\mu_0 = \mu_1 = 0$.

We consider sequences of true parameters λ_n such that $n(1-\rho_{1,n}) \to h_1$ and $\lambda_n \to \lambda_0 \in \Lambda$. Let $h = (h_1, \lambda_0')'$. Below we show that $t_n(\rho_{1,n}) \to_d J_h$, where $J_h = (\int_0^1 W_c dW)(\int_0^1 W_c^2)^{-1/2}$ when $h_1 \in R$ and $J_h = N(0,1)$ when $h_1 = \infty$.

First we consider the case in which $h_1 \in R$ and $\rho_{1,n} \to \rho_{1,0} = 1$. Define $\overline{\rho}(L) = 1 - \sum_{j=2}^k \rho_{j,0} L^j = \Pi_{j=1}^{k-1}(1-\gamma_j(\rho_0)L)$. We have $\overline{\rho}(1) = 1 - \sum_{j=2}^k \rho_{j,0} = \Pi_{j=1}^{k-1}(1-\gamma_j(\rho_0)) \neq 0$, where the inequality holds because $|\gamma_1(\rho_0)| \leq \cdots \leq |\gamma_{k-1}(\rho_0)| \leq 1-\delta$ for some $\delta > 0$. Furthermore, all roots of $\overline{\rho}(z)$ are outside the unit circle, as assumed in Theorem 2 of Hansen (1999). Following the proof of Theorem 2 of Hansen (1999), the limit distribution of $t_n(\rho_{1,n})$ is J_h when $c = h_1/\overline{\rho}(1) \in R$. The proof of Theorem 2 of Hansen (1999) is for $\rho_{1,n} = 1 + C/n$ for some $C \in R$ and for fixed $(\rho_2, \ldots, \rho_k)'$. The proof can be adjusted to apply here by (i) replacing C with $h_{1,n}$ with $h_{1,n} \to h_1 \in R$ and (ii) replacing $(\rho_2, \ldots, \rho_k)'$ with $(\rho_{2,n}, \ldots, \rho_{k,n})'$, which converges to $(\rho_{2,0}, \ldots, \rho_{k,0})'$.

Next, we show that $t_n(\rho_{1,n}) \to_d N(0,1)$ when $n(1-\rho_{1,n}) \to h_1 = \infty$. This includes the stationary case, where $\rho_{1,n} \to \rho_{1,0} < 1$, and the near stationary case, where $\rho_{1,n} \to \rho_{1,0} = 1$ and $h_1 = \infty$. To this end, we rescale $X_t = (Y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-k+1})'$ by a matrix $\Gamma_n = Diag^{-1/2}(Var_n(Y_{t-1}), \ldots, Var_n(\Delta Y_{t-k+1}))$ and define $\widetilde{X}_t = \Gamma_n X_t$, where $Var_n(\cdot)$ denotes the variance when the true parameter is λ_n. The rescaling of X_t is necessary because $Var_n(Y_{t-1})$ diverges when $\rho_{1,n} \to 1$, see Giraitis and Phillips (2006). Without loss of generality, we assume $\rho_{1,n} < 1$, although it could be arbitrarily close to 1. Let $\Sigma_n = Corr_n(\widetilde{X}_t, \widetilde{X}_t)$ when the true value is λ_n and $\sigma^2 = E_{F_0} U_t^2$. We have

$$n^{-1} \sum_{t=1}^n \widetilde{X}_t \widetilde{X}_t' - \Sigma_n \to_p 0 \text{ and } n^{-1/2} a_n' \sum_{t=1}^n \widetilde{X}_t U_t \to_d N(0, \sigma^2) \tag{A.94}$$

for any $k \times 1$ vector a_n such that $a_n' \Sigma_n a_n = 1$. The first result in (A.94) is established by showing L^2 convergence and the second is established using a triangular array martingale difference central limit theorem. For the case of an AR(1) process, these results hold by the arguments used to prove Lemmas 1 and 2 in Giraitis and Phillips (2006). Mikusheva (2007) extends these arguments to the case of an AR(k) model, as is considered here, see the proofs of (S10)-(S13) in the Supplemental Material to Mikusheva (2007) (which is available on the Econometric Society website). The proof relies on a key condition $n(1-|\gamma_k(\rho_n)|) \to \infty$. Now we show that this condition is implied by $n(1-\rho_{1,n}) \to \infty$. The result is obvious when $\rho_{1,n} \to \rho_{1,0} < 1$. When $\rho_{1,n} \to \rho_{1,0} = 1$, $\gamma_k(\rho_n) \in R$ for n large enough and $\gamma_k(\rho_n) \to 1$ because (i) $1-\rho_1 = \Pi_{j=1}^k(1-\gamma_j(\rho))$, (ii) $|\gamma_{k-1}(\rho)| \leq 1-\delta$ for some $\delta > 0$ and (iii) complex roots appear in pairs. Hence, $n(1-|\gamma_k(\rho_n)|) = n|1-\gamma_k(\rho_n)| \geq 2^{-(k-1)} n(1-\rho_{1,n})$ for n large enough, which implies that $n(1-|\gamma_k(\rho_n)|) \to \infty$.

Applying (A.94) with $a_n = \Sigma_n^{-1} l_1/(l_1' \Sigma_n^{-1} l_1)^{1/2}$ and $l_1 = (1, 0, \ldots, 0)' \in R^k$ and using $n^{-1}\sum_{t=1}^n U_t^2 \to_p \sigma^2$, which holds by (15) of Hansen (1999), we have $t_n(\rho_{1,n}) \to_d N(0,1)$ when $n(1-\rho_{1,n}) \to h_1 = \infty$.

Now we consider the behavior of the grid bootstrap critical value. Define $\widehat{h}_n = (n(1-\rho_{1,n}), \rho_{1,n}, \widehat{\rho}_2(\rho_{1,n}), \ldots, \widehat{\rho}_k(\rho_{1,n}), \widehat{F})'$, which corresponds to the true value for the bootstrap sample with sample size n. We have $\widehat{h}_n \to_p h$ because $\widehat{\rho}_j(\rho_{1,n}) - \rho_{j,n} \to_p 0$ for $j = 2, \ldots, k$ and $d_{2r}(\widehat{F}, F_0) \to 0$, where the convergence wrt the Mallows (1972) metric follows from Hansen (1999), which in turn Shao and Tu (1995, Section 3.1.2).

Let $J_n(x|h_n)$ denote the df of $t_n(\rho_{1,n})$, where $h_n = h_n(\lambda_n)$ and λ_n is true parameter vector. Then, $J_n(x|\widehat{h}_n)$ is the df of the bootstrap t statistic with sample size n. Let $J(x|h)$ denote the df of J_h. Define $L_n(h_n, h) = \sup_{x \in R} |J_n(x|h_n) - J(x|h)|$. For all non-random sequences $\{h_n : n \geq 1\}$ such that $h_n \to h$, $L_n(h_n, h) \to 0$ because $t_n(\rho_{1,n}) \to_d J_h$ and $J(x|h)$ is continuous for all $x \in R$. (For the uniformity over x in this result, see Theorem 2.6.1 of Lehmann (1999).)

Next, we show $L_n(\widehat{h}_n, h) \to_p 0$ given that $\widehat{h}_n \to_p h$ and $L_n(h_n, h) \to 0$ for all sequences $\{h_n : n \geq 1\}$ such that $h_n \to h$. Suppose $d(\cdot, \cdot)$ is a distance function (not necessarily a metric) wrt which $d(h_n, h) \to 0$ and $d(\widehat{h}_n, h) \to_p 0$.[24] Let $B(h, \varepsilon) = \{h^* \in [0, \infty) \times \Lambda : d(h^*, h) \leq \varepsilon\}$. The claim holds because (i) $\sup_{h^* \in B(h, \varepsilon_n)} L_n(h^*, h) \to 0$ for any sequence $\{\varepsilon_n : n \geq 1\}$ such that $\varepsilon_n \to 0$ and (ii) there exists a sequence $\varepsilon_n \to 0$ such that $P(d(\widehat{h}_n, h) \leq \varepsilon_n) \to 1$.[25,26]

Using the result that $\sup_{x \in R} |J_n(x|\widehat{h}_n) - J(x|h)| \to_p 0$, we have $J_n(t_n(\rho_{1,n})|\widehat{h}_n) = J(t_n(\rho_{1,n})|h) + o_p(1) \to_d U[0,1]$. The convergence in distribution holds because for all $x \in (0,1)$, $P(J(t_n(\rho_{1,n})|h) \leq x) = P(t_n(\rho_{1,n}) \leq J^{-1}(x|h)) \to J(J^{-1}(x|h)|h) = x$, where $J^{-1}(x|h)$ is the x quantile of J_h. This implies that $P(\rho_1 \in C_{g,n}) = P(\alpha/2 \leq J_n(t_n(\rho_{1,n})|\widehat{h}_n) \leq 1-\alpha/2) \to 1-\alpha$. □

---

[24] The distance can be defined as follows. Suppose $h^* = (h_1^*, \rho_1^*, \rho_2^*, \ldots, \rho_k^*, F^*)' \in [0, \infty) \times \Lambda$ and $h = (h_1, \rho_{1,0}, \rho_{2,0}, \ldots, \rho_{k,0}, F_0)' \in [0, \infty] \times \Lambda$. When $h_1 < \infty$, let $d_1(h_1^*, h_1) = |h_1^* - h_1|$. When $h_1 = \infty$, let $d_1(h_1^*, h_1) = 1/h_1^*$. Without loss of generality, assume $h_1^* \neq 0$ when $h_1 = \infty$. The distance between h^* and h is $d(h^*, h) = d_1(h_1^*, h_1) + \sum_{j=1}^k |\rho_j^* - \rho_{j,0}| + d_{2r}(F^*, F_0)$.

[25] To see that (i) holds, let $h_n^* \in B(h, \varepsilon_n)$ be such that $L_n(h_n^*, h) \geq \sup_{h^* \in B(h, \varepsilon_n)} L_n(h^*, h) - \delta_n$ for all $n \geq 1$, for some sequence $\{\delta_n : n \geq 1\}$ such that $\delta_n \to 0$. Then, $h_n^* \to h$. Hence, $L_n(h_n^*, h) \to 0$. This implies $\sup_{h^* \in B(h, \varepsilon_n)} L_n(h^*, h) \to 0$.

[26] The proof of (ii) is as follows. For all $k \geq 1$, $P(d(\widehat{h}_n, h) \leq 1/k) \geq 1 - 1/k$ for all $n \geq N_k$ for some $N_k < \infty$ because $\widehat{h}_n \to_p h$. Define $\varepsilon_n = 1/k$ for $n \in [N_k, N_{k+1})$ for $k \geq 1$. Then, $\varepsilon_n \to 0$ as $n \to \infty$ because $N_k < \infty$ for all $k \geq 1$. In addition, $P(d(\widehat{h}_n, h) \leq \varepsilon_n) = P(d(\widehat{h}_n, h) \leq 1/k) \geq 1 - 1/k$ for $n \in [N_k, N_{k+1})$, which implies that $P(d(\widehat{h}_n, h) \leq \varepsilon_n) \to 1$ as $n \to \infty$.

A.4. Quasi-likelihood ratio confidence intervals in nonlinear regression

Next, we prove Eqs. (5.10) and (5.11) for the nonlinear regression example. Eq. (5.10) holds by Theorem 4.2 of Andrews and Cheng (2012a) (AC1) provided Assumptions A, B1–B3, C1–C5, RQ1, and RQ3 of AC1 hold.

Let $\gamma_0 = (\theta_0, \phi_0)$. Eq. (5.11) holds by Theorem 4.3 and (4.15) of AC1 provided Assumptions A, B1–B3, C1–C5, C7, C8, D1–D3, and RQ1–RQ3 of AC1 hold.

Assumptions A, B1–B3, C1–C5, C7, C8, and D1–D3 of AC1 hold by Appendix E in Andrews and Cheng (2012b). It remains to verify Assumptions RQ1–RQ3 of AC1 when $r(\theta) = \beta$ and when $r(\theta) = \pi$.

We now state and prove Assumptions RQ1–RQ3 of AC1. To state these assumptions, we need to introduce some additional notation. The function $r(\theta)$ is of the form

$$r(\theta) = \begin{bmatrix} r_1(\psi) \\ r_2(\pi) \end{bmatrix}, \tag{A.95}$$

where $r_1(\psi) \in R^{d_{r_1}}$, $d_{r_1} \geq 0$ is the number of restrictions on ψ, $r_2(\pi) \in R^{d_{r_2}}$, $d_{r_2} \geq 0$ is the number of restrictions on π, and $d_r = d_{r_1} + d_{r_2}$.

The matrix $r_\theta(\theta)$ of partial derivatives of $r(\theta)$ can be written as

$$r_\theta(\theta) = \frac{\partial}{\partial\theta'} r(\theta) = \begin{bmatrix} r_{1,\psi}(\psi) & 0_{d_{r_1}} \\ 0_{d_{r_2} \times d_\psi} & r_{2,\pi}(\pi) \end{bmatrix}, \tag{A.96}$$

where $r_{1,\psi}(\psi) = (\partial/\partial\psi')r_1(\psi) \in R^{d_{r_1} \times d_\psi}$ and $r_{2,\pi}(\pi) = (\partial/\partial\pi')r_2(\pi) \in R^{d_{r_2}}$.

In particular, if $r(\theta) = \beta$, then $r_1(\psi) = \beta$, $r_2(\pi)$ does not appear and $r_{1,\psi}(\psi) = (\partial/\partial\psi')r_1(\psi) = (1, 0'_{d_\zeta}) = e'_1$. If $r(\theta) = \pi$, then $r_1(\psi)$ and $r_{1,\psi}(\psi)$ do not appear, and $r_2(\pi) = \pi$.

By definition, $\Pi_{r,0} = \Pi_r(v_{0,2})$, where $v_{0,2} = r_2(\pi_0)$ and $\gamma_0 = (\theta_0, \phi_0) \in \Gamma$. That is, $\Pi_{r,0}$ is the set of values π that are compatible with the restrictions on π when γ_0 is the true parameter value. Hence, if $r(\theta) = \beta$, then $\Pi_{r,0} = \Pi$. If $r(\theta) = \pi$, then $\Pi_{r,0} = \pi_0$.

The quantity $\widehat{s}_n$ that appears in the definition of QLR_n of AC1 is $\widehat{s}_n = \widehat{\sigma}_n^2$ in the nonlinear regression case. Also, the quantities $J(\gamma_0)$ and $V(\gamma_0)$ that appear in Assumptions D2 and D3 of AC1 and in Assumption RQ2 below are

$$\begin{aligned} J(\gamma_0) &= E_{\phi_0} d_i(\pi_0) d_i(\pi_0)' \text{ and} \\ V(\gamma_0) &= E_{\phi_0} U_i^2 d_i(\pi_0) d_i(\pi_0)' = \sigma_0^2 J(\gamma_0), \text{ where} \\ d_i(\pi) &= \left(h\left(X_i, \pi\right), Z_i', h_\pi\left(X_i, \pi\right)'\right)'. \end{aligned} \tag{A.97}$$

Note that $J(\gamma_0)$ is the probability limit under sequences $\{(\theta_n, \phi_n) : n \geq 1\}$ such that $(\theta_n, \phi_n) \to (\theta_0, \phi_0)$, $n^{1/2}|\beta_n| \to \infty$, and $\beta_n/|\beta_n| \to \omega_0 \in \{-1, 1\}$ of the second derivative of the LS criterion function after suitable scaling by a sequence of diagonal matrices. The matrix $V(\gamma_0)$ is the asymptotic variance matrix under such sequences of the first derivative of the LS criterion function after suitable scaling by a sequence of diagonal matrices.

The probability limit of the LS criterion function under γ_0 is $Q(\theta; \gamma_0)$. We have

$$Q(\theta; \gamma_0) = E_{\phi_0} U_i^2/2 + E_{\phi_0}(\beta_0 h(X_i, \pi_0) + Z_i'\zeta_0 - \beta h(X_i, \pi) - Z_i'\zeta)^2/2, \tag{A.98}$$

where $\gamma_0 = (\beta_0, \zeta_0', \pi_0', \phi_0)'$ and E_{ϕ_0} denotes expectation when the distribution of $(X_i', Z_i', U_i)'$ is ϕ_0.

If $r(\theta)$ includes restrictions on π, i.e., $d_{r_2} > 0$, then not all values $\pi \in \Pi$ are consistent with the restriction $r_2(\pi) = v_2$. For $v_2 \in r_2(\Theta)$, the set of π values that are consistent with $r_2(\pi) = v_2$ is denoted by

$$\Pi_r(v_2) = \{\pi \in \Pi : r_2(\pi) = v_2 \text{ for some } \theta = (\psi, \pi) \in \Theta\}. \tag{A.99}$$

If $d_{r_2} = 0$, then by definition $\Pi_r(v_2) = \Pi \ \forall v_2 \in r_2(\Theta)$. In consequence, if $r(\theta) = \beta$, then $\Pi_r(v_2) = \Pi$. If $r(\theta) = \pi$, then $\Pi_r(v_2) = v_2$.

Assumptions RQ1–RQ3 of AC1 are as follows.

Assumption RQ1. (i) $r(\theta)$ is continuously differentiable on Θ.
(ii) $r_\theta(\theta)$ is full row rank $d_r \ \forall\theta \in \Theta$.
(iii) $r(\theta)$ satisfies (A.95).
(iv) $d_H(\Pi_r(v_2), \Pi_r(v_{0,2})) \to 0$ as $v_2 \to v_{0,2} \ \forall v_{0,2} \in r_2(\Theta^*)$.
(v) $Q(\psi, \pi; \gamma_0)$ is continuous in ψ at ψ_0 uniformly over $\pi \in \Pi$ (i.e., $\sup_{\pi\in\Pi} |Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)| \to 0$ as $\psi \to \psi_0$) $\forall\gamma_0 \in \Gamma$ with $\beta_0 = 0$.
(vi) $Q(\theta; \gamma_0)$ is continuous in θ at $\theta_0 \ \forall\gamma_0 \in \Gamma$ with $\beta_0 \neq 0$.
In Assumption RQ1(iv), d_H denotes the Hausdorff distance.

Assumption RQ2. (i) $V(\gamma_0) = s(\gamma_0)J(\gamma_0)$ for some non-random scalar constant $s(\gamma_0) \ \forall\gamma_0 \in \Gamma$, or (ii) $V(\gamma_0)$ and $J(\gamma_0)$ are block diagonal (possibly after reordering their rows and columns), the restrictions $r(\theta)$ only involve parameters that correspond to one block of $V(\gamma_0)$ and $J(\gamma_0)$, call them $V_{11}(\gamma_0)$ and $J_{11}(\gamma_0)$, and for this block $V_{11}(\gamma_0) = s(\gamma_0)J_{11}(\gamma_0)$ for some non-random scalar constant $s(\gamma_0) \ \forall\gamma_0 \in \Gamma$.

Assumption RQ3. The scalar statistic $\widehat{s}_n$ satisfies $\widehat{s}_n \rightarrow_p s(\gamma_0)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

Assumptions RQ1(i)–(iii) hold immediately for $r(\theta) = \beta$ and $r(\theta) = \pi$. Assumption RQ1(iv) also holds because, from above, if $r(\theta) = \beta$, then $\Pi_r(v_2) = \Pi$ and if $r(\theta) = \pi$, then $\Pi_r(v_2) = v_2$. Assumption RQ1(v) holds because when $\beta_0 = 0$ we have

$$\sup_{\pi \in \Pi} |Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)| = \sup_{\pi \in \Pi} E_{\phi_0}(Z_i'(\zeta - \zeta_0) + \beta h(X_i, \pi))^2/2 \rightarrow 0 \tag{A.100}$$

as $(\beta, \zeta) \rightarrow (0, \zeta_0)$, where the convergence uses conditions in Φ^{**}. Assumption RQ1(vi) holds because

$$Q(\theta; \gamma_0) - Q(\theta_0; \gamma_0) = E_{\phi_0}(\beta h(X_i, \pi) - \beta_0 h(X_i, \pi_0) + Z_i'(\zeta - \zeta_0))^2/2 \rightarrow 0 \tag{A.101}$$

as $\theta \rightarrow \theta_0$, where the convergence uses condition in Φ^{**}.

Assumption RQ2(i) holds with $s(\gamma_0) = \sigma_0^2$ by (A.97). Assumption RQ3 holds with $\widehat{s}_n = \widehat{\sigma}_n^2$ and $s(\gamma_0) = \sigma_0^2$ by the same argument as used to verify Assumption V2 given in Appendix E of Andrews and Cheng (2012b).

References

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete set of stochastic equations. Ann. Math. Stat. 20, 46–63.

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

Andrews, D.W.K., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69, 683–734.

Andrews, D.W.K., 2002. Generalized method of moments estimation when a parameter is on a boundary. J. Bus. Econom. Statist. 20, 530–544.

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

Andrews, D.W.K., Cheng, X., 2012b. Supplemental material to 'Estimation and inference with weak, semi-strong, and strong identification'. Econom. Suppl. Mater. 80, http://www.econometricsociety.org/ecta/Supmat/9456_miscellaneous.pdf.

Andrews, D.W.K., Cheng, X., 2013. Maximum likelihood estimation and uniform inference with sporadic identification failure. J. Econometrics 173, 36–56, 2011. Supplemental Material is available with Cowles Foundation Discussion Paper No. 1824R. Yale University.

Andrews, D.W.K., Cheng, X., 2014. GMM estimation and uniform subvector inference with possible identification failure. Econometric Theory 30, 287–333.

Andrews, D.W.K., Guggenberger, P., 2009a. Hybrid and size-corrected subsampling methods. Econometrica 77, 721–762.

Andrews, D.W.K., Guggenberger, P., 2009b. Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators. J. Econometrics 152, 19–27.

Andrews, D.W.K., Guggenberger, P., 2009c. Validity of subsampling and 'plug-in asymptotic' inference for parameters defined by moment inequalities. Econometric Theory 25, 669–709.

Andrews, D.W.K., Guggenberger, P., 2010a. Asymptotic size and a problem with subsampling and the *m* out of *n* bootstrap. Econometric Theory 26, 426–468.

Andrews, D.W.K., Guggenberger, P., 2010b. Applications of hybrid and size-corrected subsampling methods. J. Econometrics 158, 285–305.

Andrews, D.W.K., Guggenberger, P., 2012. Asymptotics for LS, GLS, and feasible GLS statistics in an AR(1) model with conditional heteroskedasticity. J. Econometrics 169, 196–210.

Andrews, D.W.K., Guggenberger, P., 2014. A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter. Rev. Econ. Stat. 96, 376–381, Supplemental material available in Cowles Foundation Discussion Paper No. 1812R. Yale University.

Andrews, D.W.K., Guggenberger, P., 2017. Asymptotic size of Kleibergen's LM and conditional LR tests for moment condition models. Econometric Theory 33, 1046–1080.

Andrews, D.W.K., Guggenberger, P., 2019. Identification- and singularity-robust inference for moment condition models. Quant. Econ. 10, 1703–1746.

Andrews, D.W.K., Marmer, V., Yu, Z., 2019. On optimal inference in the linear iv model. Quant. Econ. 10, 457–485.

Andrews, I., Mikusheva, A., 2015. Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models. Quant. Econ. 6, 123–152.

Andrews, I., Mikusheva, A., 2016. Conditional inference with a functional nuisance parameter. Econometrica 84, 1571–1612.

Andrews, D.W.K., Moreira, M.J., Stock, J.H., 2004. Optimal invariant similar tests for instrumental variables regression. Cowles Foundation Discussion Paper No. 1476. Yale University.

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

Andrews, D.W.K., Moreira, M.J., Stock, J.H., 2008. Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments. J. Econometrics 146, 241–254.

Andrews, D.W.K., Shi, X., 2013. Inference based on conditional moment inequalities. Econometrica 81, 609–666, Supplemental material available at Econometrica Supplemental Material 81, http://www.econometricsociety.org/ecta/supmat/9370_Proofs.pdf.

Andrews, D.W.K., Shi, X., 2014. Nonparametric inference based on conditional moment inequalities. J. Econometrics 179, 31–45, 2011. Appendix of proofs available as Cowles Foundation Discussion Paper No. 1840RR. Yale University.

Andrews, D.W.K., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–157.

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

Cavanagh, C., 1985. Roots local to unity. Unpublished working paper. Department of Economics, Harvard University.

Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Ann. Statist. 15, 1050–1063.

Cheng, X., 2015. Robust inference in nonlinear models with mixed identification strength. J. Econometrics 189, 207–228.

Cheng, X., Liao, Z., Shi, R., 2019. On uniform asymptotic risk of averaging GMM estimators. Quant. Econ. 10, 931–979.

Chernoff, H., 1954. On the distribution of the likelihood ratio. Ann. Math. Stat. 54, 573–578.

Cox, G., 2016. Weak identification in a class of generically identified models with an application to factor models. Unpublished working paper. Department of Economics, Yale University.

Elliott, G., 1999. Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. Internat. Econom. Rev. 40, 767–783.

Elliott, G., Stock, J.H., 2001. Confidence intervals for autoregressive coefficients near one. J. Econometrics 103, 155–181.

Feir, D., Lemieux, T., Marmer, V., 2016. Weak identification in fuzzy regression discontinuity designs. J. Bus. Econom. Statist. 34, 185–196.

Geyer, C.J., 1994. On the asymptotics of constrained M-estimation. Ann. Statist. 22, 1993–2010.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.

Giraitis, L., Phillips, P.C.B., 2006. Uniform limit theory for stationary autoregression. J. Time Series Anal. 27, 51–60, Corrigendum. J. Time Ser. Anal 27, i–ii.
Guggenberger, P., 2012. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory 28, 387–421.
Guggenberger, P., Kleibergen, F., Mavroeidis, S., 2019. A more powerful subvector anderson rubin test in linear instrumental variables regression. Quant. Econ. 10, 487–526.
Guggenberger, P., Kleibergen, F., Mavroeidis, S., Chen, L., 2012a. On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression. Econometrica 80, 2649–2666.
Guggenberger, P., Ramalho, J.J.S., Smith, R.J., 2012b. GEL statistics under weak identification. J. Econometrics 170, 331–349.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Applications. Academic Press, New York.
Han, S., McCloskey, A., 2019. Estimation and inference with a (nearly) singular Jacobian. Quant. Econ. 10, 1019–1068.
Hansen, B.E., 1999. The grid bootstrap and the autoregressive model. Rev. Econ. Stat. 81, 594–607.
Kabaila, P., 1995. The effect of model selection on confidence regions and prediction regions. Econometric Theory 11, 537–549.
Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70, 1781–1803.
Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73, 1103–1123.
Leeb, H., 2006. The distribution of a linear predictor after model selection: unconditional finite-sample distributions and asymptotic approximations. In: 2nd Lehmann Symposium—Optimality. In: Lecture Notes—Monograph Series, Institute of Mathematical Statistics, Hayward, CA.
Leeb, H., Pötscher, B.M., 2005. Model selection and inference: facts and fiction. Econometric Theory 21, 21–59.
Lehmann, E.L., 1999. Elements of Large-Sample Theory. Springer, New York.
Linton, O., Song, K., Whang, Y.-J., 2010. An improved bootstrap test of stochastic dominance. J. Econometrics 154, 186–202.
Mallows, C.L., 1972. A note on asymptotic joint normality. Ann. Math. Stat. 39, 755–771.
McCloskey, A., 2017. Bonferroni-based size-correction for nonstandard testing problems. J. Econometrics 200, 17–35.
Mikusheva, A., 2007. Uniform inference in autoregressive models. Econometrica 75, 1411–1452; 2010. Corrigendum. Econometrica 78, 1733.
Mikusheva, A., 2010. Robust confidence sets in the presence of weak instruments. J. Econometrics 157, 236–247.
Mikusheva, A., Poi, B., 2006. Tests and confidence sets with correct size when instruments are potentially weak. Stata J. 6, 335–347.
Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71, 1027–1048.
Moreira, M.J., 2009. Tests with correct size when instruments can be arbitrarily weak. J. Econometrics 152, 131–140.
Müller, U.K., Elliott, G., 2003. Tests for unit roots and the initial condition. Econometrica 71, 1269–1286.
Park, J.Y., 2002. Weak unit roots. Unpublished working paper. Department of Economics, Rice University.
Phillips, P.C.B., 1987. Toward a unified asymptotic theory for autoregression. Biometrika 74, 535–547.
Phillips, P.C.B., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. J. Econometrics 136, 115–130.
Pollard, D., 1990. Empirical processes: Theory and applications. In: NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Institute of Mathematical Statistics, Hayward, CA.
Qu, Z., 2014. Inference in DSGE models with possible weak identification. Quant. Econ. 5, 457–494.
Romano, J.P., Shaikh, A.M., 2008. Inference for identifiable parameters in partially identified econometric models. J. Statist. Inference Plann. 138, 2786–2807, (Special Issue in Honor of T. W. Anderson).
Romano, J.P., Shaikh, A.M., 2010. Inference for the identified set in partially identified econometric models. Econometrica 78, 169–211.
Romano, J.P., Shaikh, A.M., 2012. On the uniform asymptotic validity of subsampling and the bootstrap. Ann. Statist. 40, 2798–2822.
Self, S.G., Liang, K.-Y., 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Amer. Statist. Assoc. 82, 605–610.
Shao, J., Tu, D., 1995. The Jackknife and Bootstrap. Springer, New York.
Shapiro, A., 1989. Asymptotic properties of statistical estimators in stochastic programming. Ann. Statist. 17, 841–858.
Staiger, D., Stock, J.H., 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. J. Monetary Econ. 28, 435–459.
Stock, J.H., Wright, J.H., 2000. GMM with weak instruments. Econometrica 68, 1055–1096.

Please cite this article as: D.W.K. Andrews, X. Cheng and P. Guggenberger, Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.027.