GENERIC RESULTS FOR ESTABLISHING THE ASYMPTOTIC SIZE OF CONFIDENCE SETS AND TESTS

By

Donald W.K. Andrews, Xu Cheng and Patrik Guggenberger

August 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1813

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

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

Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests

Donald W. K. Andrews
Cowles Foundation
Yale University

Xu Cheng
Department of Economics
University of Pennsylvania

Patrik Guggenberger
Department of Economics
UCSD

First Version: August 31, 2009
Revised: July 28, 2011

Abstract

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 has a discontinuity in its limit distribution.

The results are illustrated in three 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 k-th 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.

1 Introduction

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 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 $Y_i = \rho Y_{i-1} + U_i$ for $i = 1, ..., n$, 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 $\delta \leq 1$. 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, 2010a) 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 (2010b) 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, 2010a,b) applies to subsampling and bootstrap methods, but not to other methods of constructing critical values. It only yields asymptotic size results in cases where the test or CS has correct asymptotic size. It does not yield an explicit formula for the asymptotic size.

The approach of Andrews and Guggenberger (2009a, 2010a) applies to subsampling, hybrid, and fixed critical values, but is not designed for other methods. In this paper, we take this approach and generalize it so that it applies to a wide variety of cases, including any test statistic and any type of critical value. This approach is found to be quite flexible and easy to apply. It establishes asymptotic size whether or not 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 heteroskedasticity-robust version of the conditional likelihood ratio (CLR) test of Moreira (2003) for the linear IV regression model with included exogenous variables. This test is designed to be robust to weak identification and heteroskedasticity. In addition, it has approximately optimal power in a certain sense in the weak and strong IV scenarios with homoskedastic normal errors. 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, Moreira, and Stock (2004, Section 9), Kleibergen (2005), Guggenberger, Ramalho, and Smith (2008), Kleibergen and Mavroeidis (2009), and Guggenberger (2011).

The second 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 third 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 (2010a).

The method of this paper also has been used in Andrews and Cheng (2010a,b,c) to establish the asymptotic size of a variety of CS's based on t statistics, Wald statistics, and QLR statistics in models that exhibit lack of identification at some points in the parameter space.

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 (2010a,b) and Linton, Song, and Whang (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), and Andrews and Guggenberger (2011); for weak identification due to weak IV's, Staiger and Stock (1997), Stock and Wright (2000), Moreira (2003), Kleibergen (2005), and Guggenberger, Ramalho, and Smith (2008); for weak identification in other models, Cheng (2008), Andrews and Cheng (2010a,b,c), and I. Andrews and Mikusheva (2011); for parameters near a boundary, Chernoff (1954), Self and Liang (1987), Shapiro (1989), Geyer (1994), Andrews (1999, 2001, 2002), and Andrews and Guggenberger (2010b); for post-model selection inference, Kabaila (1995), Leeb and Pötscher (2005), Leeb (2006), and Andrews and Guggenberger (2009a,b).

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 $\mathcal{R}$. Let $CP_n(\lambda)$ denote the coverage probability of CS_n under λ. The *exact size* and *asymptotic size* of CS_n are denoted by

$$ExSz_n = \inf_{\lambda \in \Lambda} CP_n(\lambda) \text{ and } AsySz = \liminf_{n \to \infty} ExSz_n, \tag{2.1}$$

respectively.

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

$$\inf_{\lambda \in \Lambda} CP_n(\lambda) = \sup_{\lambda \in \Lambda} CP_n(\lambda). \tag{2.2}$$

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

$$\liminf_{n \to \infty} \inf_{\lambda \in \Lambda} CP_n(\lambda) = \limsup_{n \to \infty} \sup_{\lambda \in \Lambda} CP_n(\lambda). \tag{2.3}$$

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

$$AsyMaxCP = \limsup_{n \to \infty} \sup_{\lambda \in \Lambda} CP_n(\lambda). \tag{2.4}$$

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

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

We use the following assumptions:

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

$$CP_{p_n}(\lambda_{p_n}) \to [CP^-(h), CP^+(h)] \tag{2.5}$$

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

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 (2.5) holds.[2]

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 $CP \in [0,1]$.

The typical forms of h and the index set H are given in Assumption B below.

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

[1] 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 $\{\lambda_n \in \Lambda : n \geq 1\}$, there exists a subsequence $\{p_n\}$ of $\{n\}$ for which (2.5) holds.

The flawed re-indexing argument goes as follows: Let $\{\lambda_{w_n} : n \geq 1\}$ be an arbitrary subsequence of $\{\lambda_n \in \Lambda : n \geq 1\}$. 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 $\{\lambda_n^* \in \Lambda : n \geq 1\}$ 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 $\{\lambda_{w_n} : n \geq 1\}$ 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.

[2] 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 $\lim_{n\to\infty} CP_{w_n}(\lambda_{w_n}) \in [CP^-(h), CP^+(h)]$.

case does not arise. Assumption C2 is analogous. Clearly, a sufficient condition for Assumptions C1 and C2 is $CP^-(h) = CP^+(h)\ \forall h \in H$.

Assumption S (where S stands for similar) 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.

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, $AsySz \in [\inf_{h\in H} CP^-(h), \inf_{h\in H} CP^+(h)]$ *and* $AsyMaxCP \in [\sup_{h\in H} CP^-(h), \sup_{h\in H} CP^+(h)]$.

(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 provide 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 (2011).

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.[3] 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

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

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 (2.5) holds for some $h \in H$. Hence,

$$AsySz = \lim_{n \to \infty} CP_{p_n}(\lambda_{p_n}) \geq CP^-(h) \geq \inf_{h \in H} CP^-(h). \tag{2.6}$$

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

$$AsySz \leq CP^+(h) \ \forall h \in H. \tag{2.7}$$

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

$$AsySz = \liminf_{n \to \infty} \inf_{\lambda \in \Lambda} CP_n(\lambda) \leq \liminf_{n \to \infty} \inf_{\lambda \in \Lambda} CP_{p_n}(\lambda) \leq \liminf_{n \to \infty} CP_{p_n}(\lambda_{p_n}) \leq CP^+(h). \tag{2.8}$$

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

$$\inf_{h \in H} CP^-(h) = CP^-(h_L) = CP^+(h_L) \geq \inf_{h \in H} CP^+(h) \tag{2.9}$$

by Assumption C1, and $\inf_{h \in H} CP^-(h) \leq \inf_{h \in H} CP^+(h)$ because $CP^-(h) \leq CP^+(h) \ \forall h \in H$ by Assumption A2. 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 $\inf_{h \in H} CP^-(h) = \sup_{h \in H} CP^+(h) = CP$. $\square$

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 $h_n(\lambda) = (h_{n,1}(\lambda), ..., h_{n,J}(\lambda), h_{n,J+1}(\lambda))'$, $h_{n,j}(\lambda) \in R\ \forall j \leq J$, and $h_{n,J+1}(\lambda) \in \mathcal{T}$ for some compact pseudo-metric space $\mathcal{T}$.[4] 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,b).

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.

For example, Andrews and Cheng (2010a,b,c) 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 pseudo-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 $1_{d_\beta} = (1, ..., 1)' \in R^{d_\beta}$. (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 $h_n(\lambda) = (n^r\gamma_1, m_n^r\gamma_1, \gamma_2')'$ (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.

[4] For notational simplicity, we stack J real-valued quantities and one $\mathcal{T}$-valued quantity into the vector $h_n(\lambda)$.

Define

$$H = \{h \in (R \cup \{\pm\infty\})^J \times \mathcal{T} : h_{p_n}(\lambda_{p_n}) \to h \text{ for some subsequence } \{p_n\}$$
$$\text{of } \{n\} \text{ and some sequence } \{\lambda_{p_n} \in \Lambda : n \geq 1\}\}. \tag{2.10}$$

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 below.

The parameter λ and function $h_n(\lambda)$ in Assumption B typically are of the following form: (i) For all $\lambda \in \Lambda$, $\lambda = (\lambda_1, ..., \lambda_q, \lambda_{q+1})'$, where $\lambda_j \in R$ $\forall j \leq q$ and λ_{q+1} belongs to some infinite-dimensional pseudo-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) $h_n(\lambda) = (h_{n,1}(\lambda), ..., h_{n,J}(\lambda), h_{n,J+1}(\lambda))'$ for $\lambda \in \Lambda$ is of the form:

$$h_{n,j}(\lambda) = \begin{cases} d_{n,j}\lambda_j & \text{for } j = 1, ..., J_R \\ m_j(\lambda) & \text{for } j = J_R + 1, ..., J + 1, \end{cases} \tag{2.11}$$

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

In fact, often, one has $J_R = 1$, $m_j(\lambda) = \lambda_j$ 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 $j = 1, ..., J_R$ and $m_j(\lambda_n) \to h_j$ as $n \to \infty$ for $j = J_R + 1, ..., J + 1$ for some sequence of parameters $\{\lambda_n \in \Lambda : n \geq 1\}$. For example, in an AR(1) model with AR(1) parameter ρ and i.i.d. innovations with distribution F, one can take $q = J_R = J = 1$, $\lambda_1 = 1 - \rho$, $\lambda_2 = F$, $d_{n,1} = n$, and $h_n(\lambda) = n\lambda_1$.

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.[5]

The function $h_n(\lambda)$ in (2.11) 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

[5] 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 $\beta > 0$.

exhibits a discontinuity at any $\lambda \in \Lambda$ for which (a) $\lambda_j = 0$ for all $j \leq J_R$, or alternatively, (b) $\lambda_j = 0$ for some $j \leq J_R$, as in Cheng (2008). To see this, suppose $J_R = 1$, $\lambda \in \Lambda$ 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 (2.11) can be reformulated to allow for discontinuities when $\lambda_j = \lambda_j^0$, rather than $\lambda_j = 0$, for all $j \leq J_R$ or for some $j \leq J_R$. To do so, one takes $h_{n,j}(\lambda) = d_{n,j}(\lambda_j - \lambda_j^0)$ $\forall j \leq J_R$.

The function $h_n(\lambda)$ in (2.11) also can be reformulated to allow for discontinuities at J_R different values of a single parameter λ_k for some $k \leq q$, e.g., at the values in $\{\lambda_{k,1}^0, ..., \lambda_{k,J_R}^0\}$ (rather than a single discontinuity at multiple parameters $\{\lambda_1, ..., \lambda_k\}$). In this case, one takes $h_{n,j}(\lambda) = d_{n,j}(\lambda_k - \lambda_{k,j}^0)$ for $j = 1, ..., J_R$. The function $h_n(\lambda)$ can be reformulated to allow for multiple discontinuities at each parameter value in $\{\lambda_{k_1}, ..., \lambda_{k_L}\}$, where $k_\ell \leq q$ $\forall \ell \leq L$, e.g., at the values in $\{\lambda_{k_\ell,1}^0, ..., \lambda_{k_\ell,J_{R,\ell}}^0\}$ $\forall \ell \leq L$. In this case, one takes $h_{n,j}(\lambda) = d_{n,j}(\lambda_{k_\ell} - \lambda_{k_\ell,j-S_{\ell-1}}^0)$ for $j = S_{\ell-1}+1, ..., S_\ell$ for $\ell = 1, ..., L$, where $S_\ell = \sum_{s=0}^{\ell} J_{R,s}$, $J_{R,0} = 0$ and $J_R = \sum_{s=0}^{L} J_{R,s}$.

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 $\lambda_{p_n}^* = \lambda_{p_n}$ $\forall n \geq 1$.

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

$$H = \{h \in (R \cup \{\pm\infty\})^J \times \mathcal{T} : h_n(\lambda_n) \to h \text{ for some sequence } \{\lambda_n \in \Lambda : n \geq 1\}\}. \tag{2.12}$$

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

Assumption B2*. (i) For all $\lambda \in \Lambda$, $\lambda = (\lambda_1, ..., \lambda_q, \lambda_{q+1})'$, where $\lambda_j \in R\ \forall j \leq q$ and λ_{q+1} belongs to some pseudo-metric space.
(ii) condition (ii) given in (2.11) holds.
(iii) $m_j(\lambda)$ $(= m_j(\lambda_1, ..., \lambda_{q+1}))$ is continuous in $(\lambda_1, ..., \lambda_{J_R})$ uniformly over $\lambda \in \Lambda\ \forall j = J_R + 1, ..., J+1$.[6]
(iv) The parameter space Λ satisfies: for some $\delta > 0$ and all $\lambda = (\lambda_1, ..., \lambda_{q+1})' \in \Lambda$, $(a_1\lambda_1, ..., a_{J_R}\lambda_{J_R}, \lambda_{J_R+1}, ..., \lambda_{q+1})' \in \Lambda\ \forall a_j \in (0, 1]$ if $|\lambda_j| \leq \delta$, where $a_j = 1$ if $|\lambda_j| > \delta$ for $j \leq J_R$.

The comments given above regarding (2.11) also apply to the function $h_n(\lambda)$ in Assumption B2*.

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

The proof of Theorem 2.2 is given in Section 2.3 below.

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 $CP \in [0, 1]$.

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

The relationship among the assumptions is

$$\left.\begin{array}{l} B1^* \Rightarrow B1 \\ B2^* \Rightarrow B2 \end{array}\right\} \Rightarrow B, \quad B \Rightarrow A1 \ \& \ A2, \text{ and } \begin{array}{c} B1^* \Rightarrow S \\ B^* \Rightarrow S. \end{array} \tag{2.13}$$
$$B^* \Rightarrow B,$$

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 j-th component of $h_{w_n}(\lambda_{w_n})$ for $j = 1, ..., J+1$. 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 $\{p_{j,n}\}$,

$$h_{p_{j+1,n},j}(\lambda_{p_{j+1,n}}) \to h_j \text{ for some } h_j \in R. \tag{2.14}$$

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

$$h_{p_{j+1,n},j}(\lambda_{p_{j+1,n}}) \to h_j, \text{ where } h_j = \infty \text{ or } -\infty. \tag{2.15}$$

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

Next, we show that Assumption B implies Assumption A2. Given any $h \in H$, by the definition of H in (2.10), there exists a subsequence $\{p_n\}$ and a sequence $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ 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, $h_{p_n}(\lambda_{p_n}) \to h$, $d_{p_n,j}\lambda_{p_n,j} \to h_j$ $\forall j \leq J_R$, and $m_j(\lambda_{p_n}) \to h_j$ $\forall j = J_R + 1, ..., J + 1$ 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 $\forall n \geq N$, $|\lambda_{p_n,j}| < \delta$ $\forall j \leq J_R$ 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 $n \to \infty$ $\forall j \leq J_R$.)

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

$$\lambda_{s,j}^* = \begin{cases} (d_{p_n,j}/d_{s,j})\lambda_{p_n,j} & \text{if } |h_j| < \infty \ \& \ j \leq J_R \\ \lambda_{p_n,j} & \text{if } |h_j| = \infty \ \& \ j \leq J_R, \text{ or if } j = J_R + 1, ..., J + 1. \end{cases} \tag{2.16}$$

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, $\lambda_s^* \in \Lambda$ $\forall s \geq 1$.

For all $j \leq J_R$ 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 $s \geq p_N$, and $d_{p_n,j}\lambda_{p_n,j} \to h_j$ as $n \to \infty$ by the first paragraph of the proof. Hence, $h_{s,j}(\lambda_s^*) = d_{s,j}\lambda_s^* \to h_j$ as $s \to \infty$.

For all $j \leq J_R$ with $h_j = \infty$, we have $d_{s,j}\lambda_{s,j}^* = d_{s,j}\lambda_{p_n,j} \geq d_{p_n,j}\lambda_{p_n,j}$ $\forall s \in [p_n, p_{n+1})$ with

$s \geq p_N$ 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 $j \leq J_R$ with $h_j = -\infty$ is analogous.

Next, we consider $j = J_R + 1, ..., J + 1$. Define $\lambda^{**}_s = \lambda_{p_n}$ $\forall s \in [p_n,\, p_{n+1})$ and all $n \geq N$ and $\lambda^{**}_s = \lambda^*_s$ $\forall s \leq p_N$. For $j = J_R + 1, ..., J + 1$, $m_j(\lambda_{p_n}) \to h_j$ 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 $j \leq J$ and let ρ denote the pseudo-metric on $\mathcal{T}$ when $j = J + 1$. Now, for $j = J_R + 1, ..., J + 1$, if $|h_{j_1}| < \infty$ $\forall j_1 \leq J_R$, we have: $\forall s \in [p_n,\, p_{n+1})$ with $s \geq p_N$,

$$\begin{aligned}
&\rho(m_j(\lambda^*_s), m_j(\lambda^{**}_s)) \\
=\ &\rho(m_j((d_{p_n,1}/d_{s,1})\lambda_{p_n,1}, ..., (d_{p_n,J_R}/d_{s,J_R})\lambda_{p_n,J_R}, \lambda_{p_n,J_R+1}, ..., \lambda_{p_n,q+1}), \\
&\quad m_j(\lambda_{p_n,1}, ..., \lambda_{p_n,J_R}, \lambda_{p_n,J_R+1}, ..., \lambda_{p_n,q+1})) \\
\to\ &0 \text{ as } s \to \infty, \qquad (2.17)
\end{aligned}$$

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 $n \to \infty$ $\forall j_1 \leq J_R$. Equation (2.17) 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 $j_1 \leq J_R$, then the corresponding elements of λ^*_s equal those of λ^{**}_s and the convergence in (2.17) still holds by Assumption B2*(iii). Hence, we conclude that for $j = J_R + 1, ..., J + 1$, $m_j(\lambda^*_s) \to h_j$ as $s \to \infty$.

Replacing s by n, we conclude that $\{\lambda^*_n \in \Lambda : n \geq 1\}$ satisfies $h_n(\lambda^*_n) \to h \in H$ and $\lambda^*_{p_n} = \lambda_{p_n}$ $\forall n \geq 1$ 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 $d_X \geq 0$ is a matrix of included exogenous variables, and $Z \in R^{n \times d_Z}$ for $d_Z \geq 1$ is a matrix of IV's. Denote by u_i and X_i the i-th 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

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

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

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

Let

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

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

$$\begin{aligned} g_i &= g_i(\theta_0),\ \widehat{g} = n^{-1}\sum_{i=1}^n g_i,\ \widehat{G} = n^{-1}\sum_{i=1}^n G_i,\ \widehat{\Omega} = n^{-1}\sum_{i=1}^n g_i g_i{}' - \widehat{g}\widehat{g}', \\ \widehat{\Psi} &= \widehat{\Sigma} - \widehat{\Gamma}\widehat{\Omega}^{-1}\widehat{\Gamma},\ \widehat{\Sigma} = n^{-1}\sum_{i=1}^n Z_i^\perp Z_i^{\perp\prime}\widehat{v}_i^2 - \widehat{L}\widehat{L}',\ \widehat{\Gamma} = n^{-1}\sum_{i=1}^n Z_i^\perp Z_i^{\perp\prime}\widehat{u}_i\widehat{v}_i - \widehat{g}\widehat{L}', \\ \widehat{u} &= M_X(y_1 - y_2\theta_0)\ (= u^\perp),\ \widehat{v} = M_{\overline{Z}}y_2\ (\neq v^\perp),\ \text{and } \widehat{L} = n^{-1}\sum_{i=1}^n Z_i^\perp\widehat{v}_i. \end{aligned} \tag{3.6}$$

For notational convenience, subscripts n are omitted.[7]

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:

$$\begin{aligned} AR &= n\widehat{g}'\widehat{\Omega}^{-1}\widehat{g} \text{ and } LM = n\widehat{g}'\widehat{\Omega}^{-1/2}P_{\widehat{\Omega}^{-1/2}\widehat{D}}\widehat{\Omega}^{-1/2}\widehat{g}, \text{ where} \\ \widehat{D} &= \widehat{G} - n^{-1}\sum_{i=1}^{n}(G_i - \widehat{G})g_i{}'\widehat{\Omega}^{-1}\widehat{g}. \end{aligned} \tag{3.7}$$

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:

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

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

The heteroskedasticity-robust CLR test statistic is

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

The CLR statistic has the property that for W large it is approximately equal to the LM statistic LM.[8]

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

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

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).[9] 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.

[7] The centering of $\widehat{\Omega}$, $\widehat{\Sigma}$, and $\widehat{\Gamma}$ using $\widehat{g}\widehat{g}'$, $\widehat{L}\widehat{L}'$, and $\widehat{g}\widehat{L}'$, respectively, has no effect asymptotically under the null and under local alternatives, but it does have an effect under non-local alternatives.

[8] To see this requires some calculations, see the proof of Lemma 3.1 in the Appendix.

[9] More extensive tables of critical values are given in the Supplemental Material to Andrews, Moreira, and Stock (2006), which is available on the Econometric Society website.

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, Moreira, and Stock (2006, 2008). 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).

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

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

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 $(u_i, v_i, X_i', Z_i')'$.[10] 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.

The parameter space Λ of null distributions is

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

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

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

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 (2.10) with no $J+1$ component present.

Assumption B* holds with $RP = \alpha$ $\forall h \in H$ by the following Lemma.[11]

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* $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ *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.

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. 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:

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

[11] Here, RP is the testing analogoue of CP. See Comment 4 to Theorem 2.1.

where $\rho_1 \in [-1+\varepsilon, 1]$ for some $\varepsilon > 0$, U_t is i.i.d. with unknown distribution F, and $E_F U_t = 0$.[12] The time series $\{Y_t : t \geq 1\}$ is initialized at some fixed value $(Y_0, ..., Y_{1-k})'$.[13] Let $\rho = (\rho_1, ..., \rho_k)'$. Define a lag polynomial $a(L;\rho) = 1 - \rho_1 L - \sum_{j=2}^{k} \rho_j L^{j-1}(1-L)$ and factorize it as $a(L;\rho) = \Pi_{j=1}^{k}(1 - \gamma_j(\rho)L)$, where $|\gamma_1(\rho)| \leq ... \leq |\gamma_k(\rho)|$. 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

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

for some $\varepsilon > 0$, $\delta > 0$, $M < \infty$, and $r \geq 1$. Mallow's (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 $(\widehat{\rho}_2(\rho_1), ..., \widehat{\rho}_k(\rho_1))'$ denote the constrained LS estimator of $(\rho_2, ..., \rho_k)'$ 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). Given $(\rho_1, \widehat{\rho}_2(\rho_1), ..., \widehat{\rho}_k(\rho_1), \widehat{F})'$, bootstrap samples $\{Y_t(\rho_1) : t \leq n\}$ are simulated using (4.1) and some fixed starting values $(Y_0, ..., Y_{1-k})'$.[14] 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

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

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

[13] The model can be extended to allow for random starting values, for example, along the lines of Andrews and Guggenberger (2011). More specifically, from (4.1), Y_t can be written as $Y_t = \mu_0^* + \mu_1^* t + Y_t^*$ and $Y_t^* = \rho_1 Y_{t-1}^* + \rho_2 \Delta Y_{t-1}^* + ... + \rho_k \Delta Y_{t-k+1}^* + U_t$ for some μ_0^* and μ_1^*. Let $y_0^* = (Y_0^*, ..., Y_{1-k}^*)$ 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.

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

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 $\rho_n = (\rho_{1,n}, ..., \rho_{k,n})'$ and $\rho_0 = (\rho_{1,0}, ..., \rho_{k,0})'$. Define

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

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).[15]

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

[15] 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).

Andrews and Cheng (2010a, 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. (Andrews and Cheng (2010a) does not consider QLR-based CI's in the nonlinear regression model.)

5.1 Nonlinear Regression Model

The model is

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

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 finite-dimensional parameter $\pi \in R^{d_\pi}$. 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 are interested in QLR-based CI's for 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

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

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 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 Π ($\subset R^{d_\pi}$) is compact.

The random variables $\{(X_i', Z_i', U_i)' : i = 1, ..., n\}$ 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 $h_\pi(x, \pi) \in R^{d_\pi}$ and $h_{\pi\pi}(x, \pi) \in R^{d_\pi \times d_\pi}$ 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 Π^* ($\subset R^{d_\pi}$) is compact.[16] 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. The parameter space for the true value of ϕ is

$$\begin{aligned}
&\Phi^{**} = \{\phi \in \Phi^* : E_\phi(U_i|X_i, Z_i) = 0 \text{ a.s.},\ E_\phi(U_i^2|X_i, Z_i) = \sigma^2 > 0 \text{ a.s.},\\
&E_\phi\left(\sup_{\pi\in\Pi} ||h(X_i, \pi)||^{4+\varepsilon} + \sup_{\pi\in\Pi} ||h_\pi(X_i, \pi)||^{4+\varepsilon} + \sup_{\pi\in\Pi} ||h_{\pi\pi}(X_i, \pi)||^{2+\varepsilon}\right) \leq C,\\
&||h_{\pi\pi}(X_i, \pi_1) - h_{\pi\pi}(X_i, \pi_2)|| \leq M(X_i)||\pi_1 - \pi_2||\ \forall \pi_1, \pi_2 \in \Pi \text{ for some function}\\
&M(X_i),\ E_\phi M(X_i)^{2+\varepsilon} \leq C,\ E_\phi|U_i|^{4+\varepsilon} \leq C,\ E_\phi \|Z_i\|^{4+\varepsilon} \leq C,\\
&P_\phi(a'(h(X_i, \pi_1), h(X_i, \pi_2), Z_i) = 0) < 1,\ \forall \pi_1, \pi_2 \in \Pi \text{ with } \pi_1 \neq \pi_2,\ \forall a \in R^{d_\zeta + 2}\\
&\text{with } a \neq 0,\ \lambda_{\min}(E_\phi(h(X_i, \pi), Z_i')'(h(X_i, \pi), Z_i')) \geq \varepsilon\ \forall \pi \in \Pi, \text{ and}\\
&\lambda_{\min}(E_\phi d_i(\pi) d_i(\pi)') \geq \varepsilon\ \forall \pi \in \Pi\} \qquad (5.5)
\end{aligned}$$

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

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

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,

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

[16] 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.

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

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

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

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 $\chi^2_{1,1-\alpha}$. 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

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

5.3 Asymptotic Results

Under sequences $\{(\theta_n, \phi_n) : n \geq 1\}$ such that $\theta_n \in \Theta^*$, $\phi_n \in \Phi^{**}$, $(\theta_n, \phi_n) \to (\theta_0, \phi_0)$, $\beta_0 = 0$, 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 $r(\theta) = \pi$, $\gamma_0 = (\theta_0, \phi_0)$, 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.[18]

We assume that the distribution function of $LR_\infty(b, \pi_0, \phi_0)$ is continuous at $\chi^2_{1,1-\alpha}$ $\forall b \in R, \forall \pi_0 \in \Pi^*, \forall \phi_0 \in \Phi^{**}$.[19] 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 $\theta_n \in \Theta^*$, $\phi_n \in \Phi^{**}$, $(\theta_n, \phi_n) \to (\theta_0, \phi_0)$,

[18] The random quantity $\xi(\pi; b, \pi_0, \phi_0)$ is the limit in 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 limit (in 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 $\pi \in \Pi_{r,0}$.

[19] This assumption is stronger than needed, but it is simple. It is sufficient that the df of $LR_\infty(b, \pi_0, \phi_0)$ is continuous at $\chi^2_{1,1-\alpha}$ 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 $P(LR_\infty(b_L, \pi_L, \phi_L) < \chi^2_{1,1-\alpha}) = \inf_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} P(LR_\infty(b, \pi_0, \phi_0) < \chi^2_{1,1-\alpha})$ and $P(LR_\infty(b_U, \pi_U, \phi_U) \leq \chi^2_{1,1-\alpha}) = \sup_{b \in R, \pi_0 \in \Pi^*, \phi_0 \in \Phi^{**}} P(LR_\infty(b, \pi_0, \phi_0) \leq \chi^2_{1,1-\alpha})$.

$n^{1/2}|\beta_n| \to \infty$, 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 (2010a).

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

$$\begin{aligned} CP^-(h) &= CP^+(h) = CP(b, \pi_0, \phi_0) \text{ when } |b| < \infty, \text{ where} \\ CP(b, \pi_0, \phi_0) &= P\left(LR_\infty(b, \pi_0, \phi_0) \le \chi^2_{1,1-\alpha}\right). \end{aligned} \tag{5.12}$$

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

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

where S is a random variable with χ_1^2 distribution. Hence, Assumptions C1 and C2 also hold using the condition on the df of $LR_\infty(b, \pi_0, \phi_0)$.

Assumption B2*(i) holds with $(\lambda_1, ..., \lambda_q)' = (|\beta|, \beta/|\beta|, \zeta', \pi')'$ and $\lambda_{q+1} = \phi$. Assumption B2*(ii) holds with $h_n(\lambda)$ as above, $r = 1$, $d_{n,j} = n^{1/2}$, $J = 3 + d_\zeta + d_\pi$, $m_j(\lambda) = \lambda_{j-1}$ for $j = 2, ..., J$, $m_{J+1}(\lambda) = \phi$, $\Lambda = \{\lambda = (|\beta|, \beta/|\beta|, \zeta', \pi', \phi)' : \theta = (\beta, \zeta', \pi')' \in \Theta^*, \phi \in \Phi^{**}\}$, and $\mathcal{T} = \Phi^* \supset \Phi^{**}$, where Φ^* is a compact pseudo-metric space by assumption. Assumption B2*(iii) holds immediately given the form of $m_j(\lambda)$ for $j = 2, ..., J+1$. 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 $b_1^* \ge 0$, $b_2^* \ge 0$, and b_1^* and b_2^* not both equal to 0.

Hence, by Corollary 2.1(b), we have

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

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 (2010a), 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

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

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:

$$\begin{aligned}
&\text{(i) a Box-Cox function } h(x,\pi) = (x^\pi - 1)/\pi \text{ and} \\
&\text{(ii) a logistic function } h(x,\pi) = x(1+\exp(-(x-\pi)))^{-1}. \qquad (5.16)
\end{aligned}$$

In both cases, $\{(Z_i^*, X_i, U_i) : i = 1, ..., n\}$ 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 $X_i^* \sim N(3,1)$, $\text{Corr}(Z_i^*, X_i^*) = 0.5$, 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 $\{1.50, 1.75, ..., 3.5\}$. The optimization space Π for π is $[1,4]$.

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 $\{4.5, 4.6, ..., 5.5\}$. 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 Figure 1), with a grid of 0.1 for b between 0 and

Table I. 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

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. Figure 1 plots the asymptotic coverage probability as a function of b.

Table 1 and Figure 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 in the Box-Cox and logistic models are .074, .049, .209, and .085, respectively.

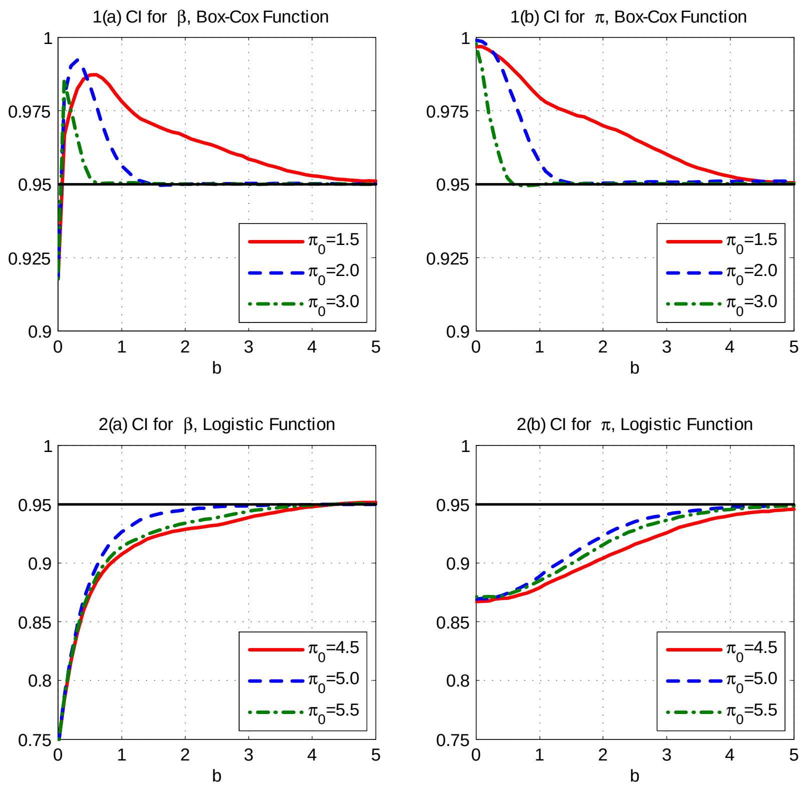

Figure 1. Asymptotic Coverage Probabilities of Standard QLR CI's for β and π in the Nonlinear Regression Model.

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:

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

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

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

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

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

where $(b, 0'_{d_\zeta})' \in R^{1+d_\zeta}$.[20]

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

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

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.

[20]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 $\pi^*(b, \pi_0, \phi_0) = \arg\min_{\pi \in \Pi} \xi(\pi; b, \pi_0, \phi_0)$.

References

ANDERSON, T. W. and RUBIN, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations", *Annals of Mathematical Statistics,* **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", *Journal of Business and Economic Statistics,* **20**, 530–544.

ANDREWS, D. W. K. and CHENG, X. (2010a), "Estimation and Inference with Weak, Semi-Strong, and Strong Identification" (Cowles Foundation Discussion Paper No. 1773R, Yale University).

ANDREWS, D. W. K. and CHENG, X. (2010b), "Maximum Likelihood Estimation and Inference with Weak, Semi-Strong, and Strong Identification", (Unpublished working paper, Cowles Foundation, Yale University).

ANDREWS, D. W. K. and CHENG, X. (2010c), "Generalized Method of Moments Estimation and Inference with Weak, Semi-Strong, and Strong Identification" (Unpublished working paper, Cowles Foundation, Yale University).

ANDREWS, D. W. K. and GUGGENBERGER, P. (2009a), "Hybrid and Size-Corrected Subsampling Methods", *Econometrica,* **77**, 721–762.

ANDREWS, D. W. K. and GUGGENBERGER, P. (2009b), "Incorrect Asymptotic Size of Subsampling Procedures Based on Post-Consistent Model Selection Estimators", *Journal of Econometrics,* **152**, 19–27.

ANDREWS, D. W. K. and 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. and 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. and GUGGENBERGER, P. (2010b), "Applications of Hybrid and Size-Corrected Subsampling Methods", *Journal of Econometrics,* **158**, 285–305.

ANDREWS, D. W. K. and GUGGENBERGER, P. (2011), "Asymptotics for LS, GLS, and Feasible GLS Statistics in an AR(1) Model with Conditional Heteroskedasticity", *Journal of Econometrics,* forthcoming.

ANDREWS, D. W. K., MOREIRA, M. J., and 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., and 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., and STOCK, J. H. (2008), "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments", *Journal of Econometrics,* **146,** 241–254.

ANDREWS, D. W. K. and SHI, X. (2010a), "Inference Based on Conditional Moment Inequalities" (Cowles Foundation Discussion Paper No. 1761R, Yale University).

ANDREWS, D. W. K. and SHI, X. (2010b), "Nonparametric Inference Based on Conditional Moment Inequalities" (Unpublished working paper, Cowles Foundation, Yale University).

ANDREWS, D. W. K. and SOARES, G. (2010), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica,* **78**, 119–157.

ANDREWS, I. and MIKUSHEVA, A. (2011): "Maximum Likelihood Inference in Weakly Identified DSGE models" (Unpublished working paper, Department of Economics, MIT).

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., and WEI, C. Z., (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *Annals of Statistics,* **15**, 1050–1063.

CHENG, X. (2008), "Robust Confidence Intervals in Nonlinear Regression under Weak Identification" (Unpublished working paper, Department of Economics, University of Pennsylvania).

CHERNOFF, H. (1954), "On the Distribution of the Likelihood Ratio", *Annals of Mathematical Statistics,* **54**, 573–578.

GEYER, C. J. (1994), "On the Asymptotics of Constrained M-Estimation", *Annals of Statistics,* **22**, 1993–2010.

GIRAITIS, L. and PHILLIPS, P. C. B. (2006), "Uniform Limit Theory for Stationary Autoregression", *Journal of Time Series Analysis,* **27**, 51–60. Corrigendum, **27**, pp. i–ii.

GUGGENBERGER, P. (2011), "On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Assumption", *Econometric Theory,* **27**, forthcoming.

GUGGENBERGER, P., RAMALHO, J. J. S. and SMITH, R. J. (2008), "GEL Statistics under Weak Identification" (Unpublished working paper, Department of Economics, Cambridge University).

HALL, P. and HEYDE, C. C. (1980), *Martingale Limit Theory and Its Applications* (New York: Academic Press).

HANSEN, B. E. (1999), "The Grid Bootstrap and the Autoregressive Model", *Review of Economics and Statistics,* **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.

KLEIBERGEN, F. and MAVROEIDIS, S. (2011), "Inference on Subsets of Parameters in Linear IV without Assuming Identification" (Unpublished working paper, Brown University).

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

LEEB, H. and PÖTSCHER, B. M. (2005): "Model Selection and Inference: Facts and Fiction", *Econometric Theory,* **21**, 21–59.

LINTON, O., SONG, K., and WHANG, Y.-J., (2010), "An Improved Bootstrap Test of Stochastic Dominance", *Journal of Econometrics,* **154**, 186–202.

MALLOWS, C. L. (1972), "A Note on Asymptotic Joint Normality", *Annals of Mathematical Statistics,* **39**, 755–771.

MIKUSHEVA, A. (2007), "Uniform Inference in Autoregressive Models", *Econometrica,* **75**, 1411–1452.

MIKUSHEVA, A. and POI, B. (2006), "Tests and Confidence Sets with Correct Size when Instruments Are Potentially Weak", *Stata Journal,* **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", *Journal of Econometrics,* **152**, 131–140.

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.

POLLARD, D. (1990), *Empirical Processes: Theory and Applications,* NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. (Hayward, CA: Institute of Mathematical Statistics).

ROMANO, J. P. and SHAIKH, A. M. (2008), "Inference for Identifiable Parameters in Partially Identified Econometric Models", *Journal of Statistical Inference and Planning* (*Special Issue in Honor of T. W. Anderson*), **138**, 2786–2807.

ROMANO, J. P. and SHAIKH, A. M. (2010a), "Inference for the Identified Set in Partially Identified Econometric Models", *Econometrica,* **78**, 169–211.

ROMANO, J. P. and SHAIKH, A. M. (2010b), "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap" (Unpublished working paper, Department of Economics, University of Chicago).

SELF, S. G. and LIANG, K.-Y. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions", *Journal of the American Statistical Association,* **82**, 605–610.

SHAPIRO, A. (1989), "Asymptotic Properties of Statistical Estimators in Stochastic Programming", *Annals of Statistics,* **17**, 841–858.

STAIGER, D. and 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", *Journal of Monetary Economics,* **28**, 435–459.

STOCK, J. H. and WRIGHT, J. H. (2000), "GMM with Weak Instruments", *Econometrica,* **68**, 1055–1096.

Appendix

for

Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests

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

Xu Cheng
Department of Economics
University of Pennsylvania

Patrik Guggenberger
Department of Economics
UCSD

August 2009
Revised: July 2011

6 Appendix

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

6.1 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 $h_n(\lambda_n) \to h = (h_1, h_2, h_{31,}..., h_{34}) \in H$. All results stated below are "under $\{\lambda_n\}$." We let E_{λ_n} denote expectation under λ_n. Define

$$\begin{aligned} h_1 &= \lim_{n\to\infty} n^{1/2}||\pi_n||,\ h_2 = \lim_{n\to\infty} \pi_n/||\pi_n||,\ h_{31} = \lim_{n\to\infty} E_{\lambda_n} Z_i^* Z_i^{*\prime}, \\ h_{32} &= \lim_{n\to\infty} \Omega_{F_n} = \lim_{n\to\infty} E_{\lambda_n} Z_i^* Z_i^{*\prime} u_i^2,\ h_{33} = \lim_{n\to\infty} \Sigma_{F_n} = \lim_{n\to\infty} E_{\lambda_n} Z_i^* Z_i^{*\prime} v_i^2,\ \text{and} \\ h_{34} &= \lim_{n\to\infty} \Gamma_{F_n} = \lim_{n\to\infty} E_{\lambda_n} Z_i^* Z_i^{*\prime} u_i v_i. \end{aligned} \tag{6.1}$$

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

$$n^{-1} \textstyle\sum_{i=1}^n A_{1i}A_{2i}A_{3i}A_{4i} - E_{\lambda_n} A_{1i}A_{2i}A_{3i}A_{4i} \to_p 0, \tag{6.2}$$

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

$$\begin{aligned} & n^{-1} \textstyle\sum_{i=1}^n g_i g_i' = n^{-1} \sum_{i=1}^n Z_i^\perp Z_i^{\perp\prime} u_i^{\perp 2} = n^{-1} \sum_{i=1}^n Z_i^* Z_i^{*\prime} u_i^2 + o_p(1),\ \text{where} \\ & Z_i^\perp = Z_i - (n^{-1} \textstyle\sum_{i=1}^n Z_i X_i')(n^{-1} \sum_{i=1}^n X_i X_i')^{-1} X_i, \\ & Z_i^* = Z_i - (E_{\lambda_n} Z_i X_i')(E_{\lambda_n} X_i X_i')^{-1} X_i,\ \text{and} \\ & u_i^\perp = u_i - (n^{-1} \textstyle\sum_{i=1}^n u_i X_i')(n^{-1} \sum_{i=1}^n X_i X_i')^{-1} X_i, \end{aligned} \tag{6.3}$$

where the second equality in the first line holds by (6.2) with $A_{1i}, ..., A_{4i}$ 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, $E_{\lambda_n} X_i u_i = 0$, 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 $F = F_n$ and $E_{F_n} = E_{\lambda_n}$.

By Lyapunov's triangular array central limit theorem (CLT), we obtain

$$\begin{aligned} n^{1/2}\widehat{g} &= n^{-1/2}\sum_{i=1}^n Z_i^{\perp}u_i^{\perp} = n^{-1/2}Z'M_X u = n^{-1/2}\sum_{i=1}^n Z_i^{\perp}u_i \\ &= n^{-1/2}\sum_{i=1}^n Z_i^{*}u_i + o_p(1) \rightarrow_d N_h \sim N(0, h_{32}), \end{aligned} \tag{6.4}$$

where the fourth equality holds using (6.2) and some calculations and the convergence uses $E_{\lambda_n}\overline{Z}_i u_i = 0$, the moment conditions in Λ, and (6.1).

Equations (6.1)-(6.4) give

$$\widehat{\Omega} = n^{-1}\sum_{i=1}^n g_i g_i' - \widehat{g}\widehat{g}' = n^{-1}\sum_{i=1}^n Z_i^{*}Z_i^{*\prime}u_i^2 + o_p(1) \rightarrow_p h_{32}. \tag{6.5}$$

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

$$\begin{aligned} \widehat{v} &= M_{\overline{Z}}y_2 = M_{\overline{Z}}v,\ \widehat{v}_i = v_i - (n^{-1}\sum_{i=1}^n v_i\overline{Z}_i)(n^{-1}\sum_{i=1}^n \overline{Z}_i\overline{Z}_i')^{-1}\overline{Z}_i, \\ \widehat{u} &= M_X(y_1 - y_2\theta_0) = M_X u,\ \text{and}\ \widehat{u}_i = u_i - (n^{-1}\sum_{i=1}^n u_i X_i)(n^{-1}\sum_{i=1}^n X_i X_i')^{-1}X_i. \end{aligned} \tag{6.6}$$

This gives

$$\begin{aligned} \widehat{L} &= n^{-1}\sum_{i=1}^n Z_i^{\perp}\widehat{v}_i = n^{-1}Z'M_X M_{\overline{Z}}y_2 = n^{-1}Z'M_{\overline{Z}}v \rightarrow_p 0, \\ \widehat{\Sigma} &= n^{-1}\sum_{i=1}^n Z_i^{\perp}Z_i^{\perp\prime}\widehat{v}_i^2 - \widehat{L}\widehat{L}' = n^{-1}\sum_{i=1}^n Z_i^{*}Z_i^{*\prime}v_i^2 + o_p(1) \rightarrow_p h_{33},\ \text{and} \\ \widehat{\Gamma} &= n^{-1}\sum_{i=1}^n Z_i^{\perp}Z_i^{\perp\prime}\widehat{u}_i\widehat{v}_i - \widehat{L}\widehat{g}' = n^{-1}\sum_{i=1}^n Z_i^{*}Z_i^{*\prime}u_i v_i + o_p(1) \rightarrow_p h_{34}, \end{aligned} \tag{6.7}$$

where the convergence in the first row holds by (6.1), (6.2), and $E_{\lambda_n}\overline{Z}_i v_i = 0$, in the second and third rows the second equalities use the result of the first row for $\widehat{L}$, (6.2), (the second and third lines of) (6.3), (6.4), (6.6), and some calculations, and the convergence in the second and third rows holds by (6.1) and (6.2).

Equations (6.5) and (6.7) and the condition $\lambda_{\min}(\Omega_F) \geq \delta > 0$ in Λ give

$$\widehat{\Psi} = \widehat{\Sigma} - \widehat{\Gamma}\widehat{\Omega}^{-1}\widehat{\Gamma} \rightarrow_p h_{33} - h_{34}h_{32}^{-1}h_{34} \equiv \Psi_h. \tag{6.8}$$

Case $h_1 < \infty$: Using the results just established, we now prove the result of the Lemma for the case $h_1 < \infty$. We have

$$n^{-1}\sum_{i=1}^n G_i g_i' = n^{-1}\sum_{i=1}^n Z_i^{\perp}Z_i^{\perp\prime}(\pi_n' Z_i^{\perp} + v_i^{\perp})u_i^{\perp} = n^{-1}\sum_{i=1}^n Z_i^{*}Z_i^{*\prime}v_i u_i + o_p(1) \rightarrow_p h_{34}, \tag{6.9}$$

where the first equality uses $y_{2i}^{\perp} = Z_i^{\perp\prime}\pi_n + v_i^{\perp}$, the second equality holds using (i) $\limsup_{n\to\infty}||\pi_n|| <$

∞, (ii) $E_{\lambda_n}\overline{Z}_i(u_i, v_i) = 0$, (iii) (6.2) 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 (6.1) and (6.2).

By (6.1), we have

$$n^{1/2}E_{\lambda_n}Z_i^*Z_i^{*\prime}\pi_n = n^{1/2}||\pi_n||E_{\lambda_n}Z_i^*Z_i^{*\prime}(\pi_n/||\pi_n||) \to h_1h_{31}h_2. \tag{6.10}$$

Using this, we obtain

$$\begin{aligned} n^{1/2}\widehat{G} &= n^{-1/2}\textstyle\sum_{i=1}^n Z_i^{\perp}y_{2i}^{\perp} = n^{-1/2}Z' M_X y_2 \\ &= n^{-1}Z' M_X Z(n^{1/2}\pi_n) + n^{-1/2}Z' M_X v \\ &= n^{-1}\textstyle\sum_{i=1}^n Z_i^*Z_i^{*\prime}(n^{1/2}\pi_n) + n^{-1/2}\sum_{i=1}^n Z_i^* v_i + o_p(1) \\ &= h_1h_{31}h_2 + n^{-1/2}\textstyle\sum_{i=1}^n Z_i^* v_i + o_p(1), \end{aligned} \tag{6.11}$$

where the fourth equality holds using $\limsup_{n\to\infty} n^{1/2}||\pi_n|| < \infty$ (since $h_1 < \infty$), (6.2), and some calculations, and the last equality holds by (6.2) and (6.10).

Using the definition of $\widehat{D}$ in (3.7), combined with (6.4), (6.5), (6.9), and (6.11) yields

$$\begin{aligned} n^{1/2}\widehat{D} &= n^{1/2}\widehat{G} - [n^{-1}\textstyle\sum_{i=1}^n G_i g_i' - \widehat{G}\widehat{g}']\widehat{\Omega}^{-1}n^{1/2}\widehat{g} \\ &= h_1h_{31}h_2 + n^{-1/2}\textstyle\sum_{i=1}^n Z_i^*v_i - h_{34}h_{32}^{-1}n^{-1/2}\sum_{i=1}^n Z_i^*u_i + o_p(1) \\ &= h_1h_{31}h_2 + [-h_{34}h_{32}^{-1} : I_{d_Z}]n^{-1/2}\textstyle\sum_{i=1}^n \begin{pmatrix} Z_i^*u_i \\ Z_i^*v_i \end{pmatrix} + o_p(1), \end{aligned} \tag{6.12}$$

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

Combining (6.4) and (6.12) gives

$$\begin{aligned} \begin{pmatrix} n^{1/2}\widehat{g} \\ n^{1/2}\widehat{D} \end{pmatrix} &= \begin{pmatrix} 0 \\ h_1h_{31}h_2 \end{pmatrix} + \begin{bmatrix} I_{d_Z} & 0 \\ -h_{34}h_{32}^{-1} & I_{d_Z} \end{bmatrix} n^{-1/2}\textstyle\sum_{i=1}^n \begin{pmatrix} Z_i^*u_i \\ Z_i^*v_i \end{pmatrix} + o_p(1) \\ \to_d \begin{pmatrix} N_h \\ D_h \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ h_1h_{31}h_2 \end{pmatrix}, \begin{bmatrix} I_{d_Z} & 0 \\ -h_{34}h_{32}^{-1} & I_{d_Z} \end{bmatrix}\begin{bmatrix} h_{32} & h_{34} \\ h_{34} & h_{33} \end{bmatrix}\begin{bmatrix} I_{d_Z} & -h_{32}^{-1}h_{34} \\ 0 & I_{d_Z} \end{bmatrix}\right) \\ &= N\left(\begin{pmatrix} 0 \\ h_1h_{31}h_2 \end{pmatrix}, \begin{bmatrix} h_{32} & 0 \\ 0 & \Psi_h \end{bmatrix}\right), \text{ where } \Psi_h = h_{33} - h_{34}h_{32}^{-1}h_{34}. \end{aligned} \tag{6.13}$$

The convergence in (6.13) holds by Lyapunov's triangular array CLT using the fact that $E_{\lambda_n}\overline{Z}_i(u_i, v_i) = 0$ implies that $E_{\lambda_n}Z_i^*u_i = E_{\lambda_n}Z_i^*v_i = 0$, the moment conditions in Λ, and (6.1). In sum, (6.13) 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, (6.13), and the continuous mapping theorem (CMT) give

$$\begin{aligned}
&(\widehat{D}'\widehat{\Omega}^{-1}\widehat{D})^{-1/2}\widehat{D}'\widehat{\Omega}^{-1}n^{1/2}\widehat{g} \to_d (D_h' h_{32}^{-1} D_h)^{-1/2} D_h' h_{32}^{-1} N_h \equiv \zeta_1 \sim N(0,1) \text{ and} \quad (6.14)\\
&LM \to_d LM_h = N_h' h_{32}^{-1/2} P_{h_{32}^{-1/2} D_h} h_{32}^{-1/2} N_h = (D_h' h_{32}^{-1} D_h)^{-1} (D_h' h_{32}^{-1} N_h)^2 = \zeta_1^2 \sim \chi_1^2,
\end{aligned}$$

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

We can write

$$AR = LM + J, \text{ where } J = n\widehat{g}'\widehat{\Omega}^{-1/2} M_{\widehat{\Omega}^{-1/2}\widehat{D}} \widehat{\Omega}^{-1/2}\widehat{g}. \quad (6.15)$$

Using (6.5), (6.8), (6.13), and the CMT, we have

$$\begin{aligned}
&J \to_d J_h = N_h' h_{32}^{-1/2} M_{h_{32}^{-1/2} D_h} h_{32}^{-1/2} N_h \sim \chi^2_{d_Z - 1} \text{ and}\\
&W = n\widehat{D}'\widehat{\Psi}^{-1}\widehat{D} \to_d W_h = D_h' \Psi_h^{-1} D_h \sim \chi^2_{d_Z}. \quad (6.16)
\end{aligned}$$

Substituting (6.15) into the definition of CLR in (3.9) and using the convergence results in (6.14) and (6.16) (which hold jointly) and the CMT, we obtain

$$CLR \to_d CLR_h = \frac{1}{2}\left(LM_h + J_h - W_h + \sqrt{(LM_h + J_h - W_h)^2 + 4LM_h W_h}\right). \quad (6.17)$$

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

$$\begin{aligned}
L_n(y) \equiv P(clr(w_n) \leq y) &= P(clr(w) + (clr(w_n) - clr(w)) \leq y)\\
&\to P(clr(w_n) \leq y) = G_L(y). \quad (6.18)
\end{aligned}$$

The distribution function $G_L(x)$ is increasing at its $(1-\alpha)$-quantile $c(1-\alpha, w)$. Therefore, by Andrews and Guggenberger (2010, 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 $clr(w)$, (6.16), and (6.17), it follows that

$$CLR - c(1-\alpha, W) \to_d CLR_h - c(1-\alpha, W_h). \tag{6.19}$$

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

$$P_{\theta_0,\lambda_n}(CLR > c(1-\alpha, W)) \to P(CLR_h > c(1-\alpha, W_h)), \tag{6.20}$$

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 (6.17) 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^2_1 and $\chi^2_{d_Z-1}$, 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 (6.20) 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

$$\begin{aligned}
&||\pi_n||^{-1}n^{-1}\textstyle\sum_{i=1}^n G_i g_i = n^{-1}\sum_{i=1}^n Z_i^\perp Z_i^{\perp\prime}((\pi_n/||\pi_n||)'Z_i^\perp + v_i^\perp/||\pi_n||)u_i^\perp = O_p(1) \text{ and}\\
&||\pi_n||^{-1}\widehat{G} = ||\pi_n||^{-1}[n^{-1}Z'M_X Z\pi_n + n^{-1}Z'M_X v] = n^{-1}\textstyle\sum_{i=1}^n Z_i^* Z_i^{*\prime}(\pi_n/||\pi_n||) + o_p(1)\\
&\to_p h_{31}h_2,
\end{aligned} \tag{6.21}$$

where the first equality in the first line uses the first equality in (6.9), the second equality in the first line uses (6.2), 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 (6.11), the second equality in the second line uses (6.2) and some calculations, and the convergence uses (6.1) and (6.2).

By the definition of $\widehat{D}$ and (6.21), we have

$$||\pi_n||^{-1}\widehat{D} = ||\pi_n||^{-1}\widehat{G} - ||\pi_n||^{-1}[n^{-1}\textstyle\sum_{i=1}^n G_i g_i' - \widehat{G}\widehat{g}']\widehat{\Omega}^{-1}\widehat{g} \to_p h_{31}h_2, \tag{6.22}$$

where $\widehat{g} = o_p(1)$ by (6.4) and $\widehat{\Omega}^{-1} = O_p(1)$ by (6.5) and the condition $\lambda_{\min}(\Omega_F) \geq \delta > 0$ in Λ.

Combining (6.4) and (6.22) gives

$$
\begin{aligned}
&(\widehat{D}'\widehat{\Omega}^{-1}\widehat{D})^{-1/2}\widehat{D}'\widehat{\Omega}^{-1}n^{1/2}\widehat{g} = (||\pi_n||^{-1}\widehat{D}'\widehat{\Omega}^{-1}||\pi_n||^{-1}\widehat{D})^{-1/2}||\pi_n||^{-1}\widehat{D}'\widehat{\Omega}^{-1}n^{1/2}\widehat{g} \\
&\to_d (h_2'h_{31}h_{32}^{-1}h_{31}h_2)^{-1/2}h_2'h_{31}h_{32}^{-1}N_h \equiv \zeta_2 \sim N(0,1) \text{ and} \\
&LM \to_d LM_h = N_h'h_{32}^{-1/2}P_{h_{32}^{-1/2}h_{31}h_2}h_{32}^{-1/2}N_h \\
&= (h_2'h_{31}h_{32}^{-1}h_{31}h_2)^{-1}(h_2'h_{31}h_{32}^{-1}N_h)^2 = \zeta_2^2 \sim \chi_1^2. \qquad (6.23)
\end{aligned}
$$

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

$$P_{\theta_0,\lambda_n}(n^{1/2}||\pi_n|| \cdot ||\pi_n||^{-1}||\widehat{D}|| > K) \to 1. \qquad (6.24)$$

This, (6.8), and $||\Psi_h|| < \infty$ (by the conditions in Λ) yield

$$P_{\theta_0,\lambda_n}(W > K) = P_{\theta_0,\lambda_n}(n\widehat{D}'\widehat{\Psi}^{-1}\widehat{D} > K) \to 1. \qquad (6.25)$$

By (6.15) and some calculations, we have

$$(AR - W)^2 + 4LM \cdot W = (LM - J + W)^2 + 4LM \cdot J. \qquad (6.26)$$

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

$$CLR = \frac{1}{2}\left(LM + J - W + \sqrt{(LM - J + W)^2 + 4LM \cdot J}\right). \qquad (6.27)$$

Using a first-order expansion of the square-root expression in (6.27) about $(LM - J + W)^2$, we obtain

$$\sqrt{(LM - J + W)^2 + 4LM \cdot J} = LM - J + W + (1/2)\xi^{-1/2}4LM \cdot J \qquad (6.28)$$

for an intermediate value ξ between $(LM - J + W)^2$ and $(LM - J + W)^2 + 4LM \cdot J$. By (6.23) and (6.25), $\xi^{-1/2}LM \cdot J \to_p 0$.

This, (6.23), (6.27), and (6.28) give

$$CLR = LM + o_p(1) \to_d \zeta_2^2 \sim \chi_1^2. \qquad (6.29)$$

Define $CLR_h(w)$ as CLR_h is defined in (6.17), 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 (6.26)-(6.29). In consequence, the $(1-\alpha)$-

quantile $c(1-\alpha, w)$ satisfies $c(1-\alpha, w) \to \chi^2_{1,1-\alpha}$ as $w \to \infty$, where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2_1 distribution. Combining this with (6.25) gives

$$c(1-\alpha, W) \to_p \chi^2_{1,1-\alpha}. \tag{6.30}$$

The result of Lemma 3.1 for the case $h_1 = \infty$ follows from (6.29), (6.30), and the definition of convergence in distribution. □

6.2 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 $|\gamma_1(\rho_0)| \leq ... \leq |\gamma_{k-1}(\rho_0)| \leq 1-\delta$ 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 $(\rho_2, ..., \rho_k)'$. 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 $(\rho_2, ..., \rho_k)'$ with $(\rho_{2,n}, ..., \rho_{k,n})'$, which converges to $(\rho_{2,0}, ..., \rho_{k,0})'$.

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 $X_t = (Y_{t-1}, \Delta Y_{t-1}, ..., \Delta Y_{t-k+1})'$ by a matrix $\Gamma_n = Diag^{-1/2}(Var_n(Y_{t-1}), ..., Var_n(\Delta Y_{t-k+1}))$ 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 $\Sigma_n = Corr(\widetilde{X}_t, \widetilde{X}_t)$ when the true value is λ_n and $\sigma^2 = E_{F_0} U_t^2$. We have

$$n^{-1} \sum_{t=1}^n \widetilde{X}_t \widetilde{X}_t' - \Sigma_n \to_p 0 \text{ and } n^{-1/2} a_n' \sum_{t=1}^n \widetilde{X}_t U_t \to_d N(0, \sigma^2) \tag{6.31}$$

for any $k \times 1$ vector a_n such that $a_n' \Sigma_n a_n = 1$. The first result in (6.31) 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 (6.31) with $a_n = \Sigma_n^{-1} l_1 / (l_1' \Sigma_n^{-1} l_1)^{1/2}$ and $l_1 = (1, 0, ..., 0)' \in R^k$ 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 $\widehat{h}_n = (n(1 - \rho_{1,n}), \rho_{1,n}, \widehat{\rho}_2(\rho_{1,n}), ..., \widehat{\rho}_k(\rho_{1,n}), \widehat{F})'$, 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 $j = 2, ..., k$ and $d_{2r}(\widehat{F}, F_0) \to 0$, where the convergence wrt the Mallows (1972) metric follows from Hansen (1999), which in turn references Shao and Tu (1995, Section 3.1.2).

Let $J_n(x|h_n)$ denote the distribution function (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 $d(\widehat{h}_n, h) \to_p 0$.[21] 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

[21] The distance can be defined as follows. Suppose $h^* = (h_1^*, \rho_1^*, \rho_2^*, ..., \rho_k^*, F^*)' \in [0, \infty) \times \Lambda$ and $h = (h_1, \rho_{1,0}, \rho_{2,0}, ..., \rho_{k,0}, F_0)' \in [0, \infty] \times \Lambda$. 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 $d(h^*, h) = d_1(h_1^*, h_1) + \sum_{j=1}^k |\rho_j^* - \rho_{j,0}| + d_{2r}(F^*, F_0)$.

(ii) there exists a sequence $\varepsilon_n \to 0$ such that $P(d(\widehat{h}_n, h) \le \varepsilon_n) \to 1$.[22,23]

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 $x \in (0,1)$, $P(J(t_n(\rho_{1,n})|h) \le x) = P(t_n(\rho_{1,n}) \le J^{-1}(x|h)) \to J(J^{-1}(x|h))|h) = x$, where $J^{-1}(x|h)$ is the x quantile of J_h. This implies that $P(\rho_1 \in C_{g,n}) = P(\alpha/2 \le J_n(t_n(\rho_{1,n})|\widehat{h}_n) \le 1 - \alpha/2) \to 1 - \alpha$. $\square$

6.3 Quasi-Likelihood Ratio Confidence Intervals in Nonlinear Regression

Next, we prove equations (5.10) and (5.11) for the nonlinear regression example. Equation (5.10) holds by Theorem 4.2 of Andrews and Cheng (2010) (AC1) provided Assumptions A, B1-B3, C1-C5, RQ1, and RQ3 of AC1 hold.

Equation (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 of AC1. 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

$$r(\theta) = \begin{bmatrix} r_1(\psi) \\ r_2(\pi) \end{bmatrix}, \tag{6.32}$$

where $r_1(\psi) \in R^{d_{r_1}}$, $d_{r_1} \ge 0$ is the number of restrictions on ψ, $r_2(\pi) \in R^{d_{r_2}}$, $d_{r_2} \ge 0$ 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

$$r_\theta(\theta) = \frac{\partial}{\partial \theta'} r(\theta) = \begin{bmatrix} r_{1,\psi}(\psi) & 0_{d_{r_1} \times d_\pi} \\ 0_{d_{r_2} \times d_\psi} & r_{2,\pi}(\pi) \end{bmatrix}, \tag{6.33}$$

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

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 $r_2(\pi) = \pi$.

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

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

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

$$\begin{aligned} J(\gamma_0) &= E_{\phi_0} d_i(\pi_0) d_i(\pi_0)' \text{ and} \\ V(\gamma_0) &= E_{\phi_0} U_i^2 d_i(\pi_0) d_i(\pi_0)' = \sigma_0^2 J(\gamma_0), \text{ where} \\ d_i(\pi) &= \left(h\left(X_i, \pi\right), Z_i', h_\pi\left(X_i, \pi\right)'\right)'. \end{aligned} \tag{6.34}$$

Note that $J(\gamma_0)$ is the probability limit under sequences $\{(\theta_n, \phi_n) : n \geq 1\}$ such that $(\theta_n, \phi_n) \rightarrow (\theta_0, \phi_0)$, $n^{1/2}|\beta_n| \rightarrow \infty$, and $\beta_n/|\beta_n| \rightarrow \omega_0 \in \{-1, 1\}$ 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

$$Q(\theta; \gamma_0) = E_{\phi_0} U_i^2/2 + E_{\phi_0}(\beta_0 h(X_i, \pi_0) + Z_i'\zeta_0 - \beta h(X_i, \pi) - Z_i'\zeta)^2/2, \tag{6.35}$$

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

$$\Pi_r(v_2) = \{\pi \in \Pi : r_2(\pi) = v_2 \text{ for some } \theta = (\psi, \pi) \in \Theta\}. \tag{6.36}$$

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 (6.32).
(iv) $d_H(\Pi_r(v_2), \Pi_r(v_{0,2})) \rightarrow 0$ as $v_2 \rightarrow v_{0,2} \; \forall v_{0,2} \in r_2(\Theta^*)$.
(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 θ_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 $\widehat{s}_n \to_p s(\gamma_0)$ 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

$$\sup_{\pi \in \Pi} |Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)| = \sup_{\pi \in \Pi} E_{\phi_0}(Z_i'(\zeta - \zeta_0) + \beta h(X_i, \pi))^2/2 \to 0 \tag{6.37}$$

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

$$Q(\theta; \gamma_0) - Q(\theta_0; \gamma_0) = E_{\phi_0}(\beta h(X_i, \pi) - \beta_0 h(X_i, \pi_0) + Z_i'(\zeta - \zeta_0))^2/2 \to 0 \tag{6.38}$$

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

Assumption RQ2(i) holds with $s(\gamma_0) = \sigma_0^2$ by (6.34). 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 AC1.

References for Appendix

ANDREWS, D. W. K. and CHENG, X. (2010), "Estimation and Inference with Weak, Semi-Strong, and Strong Identification" (Cowles Foundation Discussion Paper No. 1773, Yale University).

ANDREWS, D. W. K. and GUGGENBERGER, P. (2010), "Asymptotic Size and a Problem with Subsampling and the m Out of n Bootstrap", *Econometric Theory,* **26**, 426–468.

GIRAITIS, L. and PHILLIPS, P. C. B. (2006), "Uniform Limit Theory for Stationary Autoregression", *Journal of Time Series Analysis,* **27**, 51–60. Corrigendum, **27**, pp. i–ii.

HANSEN, B. E. (1999), "The Grid Bootstrap and the Autoregressive Model", *Review of Economics and Statistics,* **81**, 594–607.

LEHMANN, E. L. (1999), *Elements of Large-Sample Theory* (New York: Springer).

MALLOWS, C. L. (1972), "A Note on Asymptotic Joint Normality", *Annals of Mathematical Statistics,* **39**, 755–771.

MIKUSHEVA, A. (2007), "Uniform Inference in Autoregressive Models", *Econometrica,* **75**, 1411–1452.

SHAO, J., and TU, D. (1995), *The Jackknife and Bootstrap* (New York, Springer).