ESTIMATION AND INFERENCE WITH WEAK, SEMI-STRONG, AND STRONG IDENTIFICATION

By

Donald W. K. Andrews and Xu Cheng

October 2010

COWLES FOUNDATION DISCUSSION PAPER NO. 1773

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

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

Estimation and Inference with Weak, Semi-strong, and Strong Identification

Donald W. K. Andrews*
Cowles Foundation
Yale University

Xu Cheng
Department of Economics
University of Pennsylvania

First Draft: August, 2007
Revised: October 6, 2010

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0751517. The authors thank Xiaohong Chen, Sukjin Han, Yuichi Kitamura, Peter Phillips, Eric Renault, Frank Schorfheide, and Ed Vytlacil for helpful comments.

Abstract

This paper analyzes the properties of standard estimators, tests, and confidence sets (CS's) in a class of models in which the parameters are unidentified or weakly identified in some parts of the parameter space. The paper also introduces methods to make the tests and CS's robust to such identification problems. The results apply to a class of extremum estimators and corresponding tests and CS's, including maximum likelihood (ML), least squares (LS), quantile, generalized method of moments (GMM), generalized empirical likelihood (GEL), minimum distance (MD), and semi-parametric estimators. The consistency/lack-of-consistency and asymptotic distributions of the estimators are established under a full range of drifting sequences of true distributions. The asymptotic size (in a uniform sense) of standard tests and CS's is established. The results are applied to the ML estimator of an ARMA(1, 1) model and to the LS estimator of a nonlinear regression model. In companion papers the results are applied to a number of other models.

Keywords: Asymptotic size, confidence set, estimator, identification, nonlinear models, strong identification, test, weak identification.

JEL Classification Numbers: C12, C15.

1. Introduction

The literature in econometrics has shown considerable interest in issues related to identification over the last two decades (and, of course, prior to that as well). For example, research has been carried out on models with weak instruments, models with partial identification, models with and without nonparametric identification, tests with nuisance parameters that are unidentified under the null hypothesis, and the finite sample properties of statistics under lack of identification. The present paper is in this line of research, but focuses on a class of models that has not been investigated fully in the literature. It includes models with weak instruments but the focus of the paper is on other models in this class.

We consider a class of models in which lack of identification occurs in part of the parameter space. Specifically, we consider models in which the parameter θ of interest is of the form $\theta = (\beta, \zeta, \pi)$, where π is identified if and only if $\beta \neq 0$, ζ is not related to the identification of π, and $\psi = (\beta, \zeta)$ is always identified. The parameters β, ζ, and π may be scalars or vectors. This a canonical parametrization which may or may not hold in the natural parameterization of the model, but is assumed to hold after suitable reparametrization. For example, the nonlinear regression model, $Y_i = \beta h(X_i, \pi) + Z_i'\zeta + U_i$, where (Y_i, X_i, Z_i) is observed and $h(\cdot, \cdot)$ is known, is of the form just described. So are other models that depend on a nonlinear index of the form $\beta h(X_i, \pi) + Z_i'\zeta$.[1]

Suppose θ is estimated by minimizing a criterion function $Q_n(\theta)$ over a parameter space Θ. Lack of identification of π when $\beta = 0$ leads to $Q_n(\theta)$ being (relatively) flat with respect to (wrt) π when β is close to 0. For example, the LS criterion function in the nonlinear regression example, $n^{-1}\sum_{i=1}^{n}(Y_i - \beta h(X_i, \pi) - Z_i'\zeta)^2$, has first derivative wrt π equal to $-2\beta n^{-1}\sum_{i=1}^{n}(Y_i - \beta h(X_i, \pi) - Z_i'\zeta)(\partial/\partial\pi)h(X_i, \pi)$, which is close to 0 for β close to 0. Flatness of $Q_n(\theta)$ is well-known to cause numerical difficulties in practice. It also causes difficulties with standard asymptotic approximations because the second derivative matrix of $Q_n(\theta)$ is singular or near singular and standard asymptotic approximations involve the inverse of this matrix.

[1] Throughout the paper we use the term identification/lack of identification in the sense of identification by a criterion function $Q_n(\theta)$. Lack of identification by $Q_n(\theta)$ means that $Q_n(\theta)$ is flat in some directions in part of the parameter space. See Assumption A below for a precise definition. Lack of identification by the criterion function $Q_n(\theta)$ is not the same as lack of identification in the usual or strict sense of the term, although there is a close relationship. For example, with a likelihood criterion function, the former implies the latter. See Sargan (1983) for a related distinction between lack of identification in the strict sense and lack of first order identification.

This paper applies the general results to an ARMA(1, 1) model. The nonlinear regression model is treated in the Supplemental Material to this paper, Andrews and Cheng (2007) (AC1-SM).[2] Two companion papers—Andrews and Cheng (2008a,b) (hereafter AC2 and AC3, respectively) apply the results of this paper to a smooth transition threshold autoregressive (STAR) model, a smooth transition switching regression model, a nonlinear binary choice model, and a nonlinear regression model with endogenous regressors. In addition, work is underway on applications to limited dependent variable models, including probit and censored regression, with endogeneity and a linear reduced-form equation for the endogenous variable(s), see Nelson and Olson (1978), Lee (1981), Smith and Blundell (1986), Newey (1987), and Rivers and Vuong (1988), and an endogenous probit model with no exclusion restriction but a nonlinear parametric reduced-form equation for the endogenous regressor, see Dong (2009) for a related model. Han (2009) shows that, via reparametrization, a simple bivariate probit model with endogeneity falls into the class of models considered here.

Other examples covered by the results of this paper include MIDAS regressions in empirical finance, which combine data with different sampling frequencies, see Ghysels, Sinko, and Valkanov (2007), models with autoregressive distributed lags, continuous transition structural change models, continuous transition threshold autoregressive models (e.g., see Chan and Tsay (1998)), seasonal ARMA(1, 1) models (e.g., see Andrews, Liu, and Ploberger (1998)), models with correlated random coefficients (e.g., see Andrews (2001)), GARCH(p, q) models, and time series models with nonlinear deterministic time trends of the form t^{π} or $(t^{\pi}-1)/\pi$.[3]

Not all models with lack of identification at some points in the parameter space fall into the class of models considered here. The models considered here must satisfy a set of criterion function (stochastic) quadratic approximation conditions, as described in more detail below, that do not apply to some models of interest. For example, abrupt transition structural change models, (unobserved) regime switching models, and abrupt transition threshold autoregressive models are not covered by the results of the present paper, e.g., see Picard (1985), Chan (1993), Bai (1997), Hansen (2000), Liu and Shao

[2] Cheng (2008) also considers the nonlinear regression model. The treatment of this model in AC1-SM is more general than in Cheng (2008) in that it allows for a whole class of error distributions, but is less general in that it only considers a single source of potential lack of identification, i.e., a single nonlinear regressor.

[3] Nonlinear time trends can be analyzed asymptotically in the framework considered in this paper via sample size rescaling, i.e., by considering $(t/n)^{\pi}$ or $((t/n)^{\pi}-1)/\pi$, e.g., see Andrews and McDermott (1995).

(2003), Elliott and Müller (2007, 2008), and Drton (2009) for analyses of these models.

The approach of the paper is to consider a general class of extremum estimators that includes ML, LS, quantile, GMM, GEL, and MD estimators. The criterion functions considered may be smooth or non-smooth functions of θ. We place high-level conditions on the behavior of the criterion function $Q_n(\theta)$, provide a variety of more primitive sufficient conditions, and verify the latter in several examples. For example, in AC2, we provide more primitive sufficient conditions for the case where the criterion function takes the form of a sample average that is a smooth function of θ and is based on i.i.d. or stationary time series observations, which covers ML and LS estimators. These conditions are of a similar nature to standard ML regularity conditions, and indeed cover ML estimators, but allow for non-regularity in terms of a certain type of identification failure. We also provide sufficient conditions for GMM criterion functions in AC3. The high-level conditions given here have the attractive features of (i) clarifying precisely which features of the criterion function are essential for the analysis and (ii) covering a wide variety of cases simultaneously.

Given the high-level conditions, we establish the large sample properties of extremum estimators, t tests, and t CS's under lack of identification, weak identification, semi-strong identification, and strong identification, as discussed below. These large sample properties provide good approximations to the statistics' finite-sample properties under all strengths of identification, whereas standard asymptotic theory only provides good approximations under strong identification. We investigate the large sample biases of extremum estimators under weak identification. We determine the asymptotic size of standard t tests and CS's, which often deviates from their nominal size in the presence of lack of identification at some points in the parameter space.[4] In AC2, we provide corresponding results for quasi-likelihood ratio (QLR) tests and CS's. In AC3, we do likewise for Wald tests and CS's.

We introduce methods of making standard tests and CS's robust to lack of identification, i.e., to have correct asymptotic size (in a uniform sense). These methods include least-favorable (LF), type 1 robust, and type 2 robust critical values. The LF critical value is a constant that is large enough for all identification categories. The type 1 critical value is data dependent and is closely related to a method suggested in Andrews

[4]Asymptotic size is defined to be the limit of exact (i.e., finite-sample) size. For a test, exact size is the maximum rejection probability over distributions in the null hypothesis. For a CI, exact size is the minimum coverage probability over all distributions. Because exact size has uniformity built into its definition, so does asymptotic size as defined here.

(1999, Sec. 6.4; 2000, Sec. 4) for boundary problems and to the generalized moment selection critical value method used in Andrews and Soares (2010) and some other papers for inference in partially-identified models based on moment inequalities. The type 2 critical value is data dependent and is similar to that used in Andrews and Jia (2008) for inference based on moment inequalities. With type 1 and type 2 robust critical values, the idea is to use a identification-category selection procedure to determine whether β is close to the non-identification value 0 and, if so, to adjust the critical value to take account of the effect of non-identification or weak identification on the behavior of the test statistic.

We also introduce null-imposed (NI) and plug-in versions of these robust critical values. The NI version exploits the knowledge of the null hypothesis value to make the critical value smaller. The plug-in version replaces consistently estimable nuisance parameters by consistent estimators in order to make the critical value smaller. The NI and plug-in versions improve the statistic properties of the robust critical values, but often at a price in terms of computation.

The resulting identification-robust tests and CS's are ad hoc in nature and do not have any optimality properties. However, they are generally applicable and often have the advantage of computational ease. In some models with potential identification failure, procedures with explicit asymptotic optimality/admissibility properties are available. For example, see Elliott and Müller (2007, 2008) for some change-point problems.

In the models considered here, weak identification occurs when $\beta \neq 0$ but β is close to 0. As is well-known from the literature on weak instruments, the effect of β of a given magnitude on the behavior of estimators and tests depends on the sample size n. In consequence, to capture asymptotically the finite-sample behavior of estimators, tests, and CS's under near non-identification, one has to consider drifting sequences of true distributions. In the present context, one needs to consider drifting sequences in which β_n drifts to 0 at various rates and β_n drifts to non-zero values.

Interest in asymptotics with drifting sequences of parameters goes back to Neyman-Pitman drifts, which are used to approximate the power functions of tests, and contiguity results, which are used for asymptotic efficiency calculations among other things. More recently, drifting sequences of parameters have been shown to play a crucial role in the literature on weak instruments, e.g., see Staiger and Stock (1997), and the literature on the (uniform) asymptotic size properties of tests and CS's when the statistics of interest display discontinuities in their pointwise asymptotic distributions, see Andrews

Table I. Identification Categories

Category	$\{\beta_n\}$ Sequence	Identification Property of π
I(a)	$\beta_n = 0 \ \forall n \geq 1$	Unidentified
I(b)	$\beta_n \neq 0$ and $n^{1/2}\beta_n \to b \in R^{d_\beta}$ (and, hence, $\|\|\beta_n\|\| = O(n^{-1/2})$)	Weakly identified
II	$\beta_n \to 0$ and $n^{1/2}\|\|\beta_n\|\| \to \infty$	Semi-strongly identified
III	$\beta_n \to \beta_0 \neq 0$	Strongly identified

and Guggenberger (2009, 2010) and Andrews, Cheng, and Guggenberger (2009). The situation considered here is an example of the latter phenomenon. The latter papers show that to determine asymptotic size, it is both necessary and sufficient to determine the behavior of the relevant statistics under certain drifting sequences of parameters. In this paper, we use the results in those papers and consider a collection of drifting sequences of parameters/distributions that are sufficient to determine the asymptotic size of the tests and CS's considered.

Suppose the true value of the parameter is $\theta_n = (\beta_n, \zeta_n, \pi_n)$ for $n \geq 1$, where n indexes the sample size. The behavior of extremum estimators and tests in the present context depends on the magnitude of $||\beta_n||$. The asymptotic behavior of these statistics varies across the three categories of sequences $\{\beta_n : n \geq 1\}$ defined in Table I.[5]

The asymptotic results of the paper for the extremum estimator $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{\zeta}_n, \widehat{\pi}_n)$ are summarized as follows: The estimator $\widehat{\psi}_n = (\widehat{\beta}_n, \widehat{\zeta}_n)$ is $n^{1/2}$-consistent for all categories of sequences $\{\beta_n\}$. The estimator $\widehat{\pi}_n$ is inconsistent for Category I sequences and consistent for Categories II and III. The asymptotic distribution of $n^{1/2}(\widehat{\psi}_n - \psi_n)$ $(= n^{1/2}((\widehat{\beta}_n, \widehat{\zeta}_n) - (\beta_n, \zeta_n)))$ is a functional of a Gaussian process with a mean that is (typically) non-zero for Category I sequences (due to the inconsistency of $\widehat{\pi}_n$) and is normal with mean zero for Categories II and III. The asymptotic distribution of $\widehat{\pi}_n$ is a functional of the same Gaussian process for Category I sequences. These estimation results permit the calculation of the asymptotic biases of $(\widehat{\beta}_n, \widehat{\zeta}_n, \widehat{\pi}_n)$ for Category I sequences as a function of the strength of identification. The asymptotic distribution of $n^{1/2}||\beta_n||(\widehat{\pi}_n - \pi_n)$ is normal with mean zero for Category II sequences. The asymptotic distribution of $n^{1/2}(\widehat{\pi}_n - \pi_n)$ is normal with mean zero for Category III sequences.

[5] Hahn and Kuersteiner (2002) and Antoine and Renault (2009, 2010) refer to sequences in our *semi-strong* category as *nearly weak*. For this paper at least, we prefer our terminology because estimators are consistent and asymptotically normal under semi-strong sequences, just as under sequences in the strong category. The only difference is that their rate of convergence is slower.

Similarly, the asymptotic results for tests and CS's vary over the three categories. For Category I sequences, standard tests and CS's have asymptotic rejection/coverage probabilities that may differ, sometimes substantially, from their nominal level. In consequence, the asymptotic size of standard tests and CS's often is substantially different from the desired nominal size. For Category II and III sequences, standard tests and CS's have the desired asymptotic rejection/coverage probability properties. For hypotheses or CS's that involve π, their power/non-coverage properties are standard for Category II and III sequences.

The results of the paper are applied to the ARMA(1, 1) model, which is a workhorse model in applied time series analysis. It is been known for many years that common moving average (MA) and autoregressive (AR) roots leads to identification failure in the ARMA(1, 1) model in the important scenario where the series is white noise, see Ansley and Newbold (1980). Results for testing the null hypothesis of white noise in an ARMA(1, 1) model have been provided by Hannan (1982) and Andrews and Ploberger (1996). However, no papers provide an asymptotic analysis of standard estimators, CI's, or tests for any other null hypothesis (such as tests concerning the AR or MA parameter) that deal with the identification issue. We do so in this paper. We also introduce identification robust CI's and provide extensive numerical results concerning the asymptotic and finite-sample properties of a variety of estimators and CI's.

The results for the ARMA(1, 1) model are summarized as follows. The distributions of the maximum likelihood (ML) estimators of the MA and AR parameters are greatly effected by lack of identification and weak identification, both asymptotically and in finite samples. Their distributions are bi-modal, biased for non-zero true values, and far from the standard normal distribution. The asymptotic distributions for the MA and AR parameter estimators are the same under weak identification. The asymptotic approximations to the finite-sample distributions are remarkably good.

Standard t CI's are found to have asymptotic and finite-sample sizes that are very poor—less than 0.60 for nominal 95% CI's concerning the MA and AR parameters. Standard CI's based on the QLR statistic and a χ^2 critical value, on the other hand, have asymptotic and finite-sample sizes that are not correct, but are far superior to those of standard $|t|$ CI's. Their asymptotic size is 0.933 for nominal 95% CI's and their finite-sample sizes are close to this. The asymptotic approximations for the standard t and QLR CI's work very well.

The nominal 95% robust CI's have asymptotic and finite-sample size that are equal to, and close to, 0.95, respectively. This is true even for the robust CI's based on the t statistic. The best robust CI in terms of false coverage probabilities is a type 2 NI robust CI based on the QLR statistic. The asymptotic approximations for the robust CI's are found to work very well.

Next, we discuss the literature that is related to this paper. Cheng (2008) considers a nonlinear regression model with multiple nonlinear regressors and, hence, multiple sources of lack of identification. In contrast, the present paper only considers a single source of lack of identification (based on the magnitude of the true value of $||\beta||$), which translates into a single nonlinear regressor in the nonlinear regression example. On the other hand, the present paper covers a much wider variety of models than does Cheng (2008).

In the models considered in this paper, a test of $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$, is a test for which π is a nuisance parameter that is unidentified under the null hypothesis. Testing problems of this type have been considered in the literature, see Davies (1977, 1987), Andrews and Ploberger (1994), and Hansen (1996).

In contrast, the hypotheses considered in this paper are of a more general type. To obtain asymptotic size results for CS's for β, as is done here, one needs to consider drifting sequences of null hypotheses of the form $H_0 : \beta = \beta_n^*$ for $n \geq 1$. Such testing problems are not considered in the literature referenced above. Furthermore, here we consider a full range of nonlinear hypotheses concerning (β, ζ, π)—only special cases are of the type $H_0 : \beta = 0$. For example, when the null hypothesis concerns ζ, then π is a nuisance parameter that is identified in part of the null hypothesis and unidentified in another part. If the null hypothesis involves all three parameters (β, ζ, π), then the identification scenario is substantially more complicated than when H_0 is $\beta = 0$.

The weak instrumental variable (IV) literature, e.g., see Nelson and Startz (1990), Dufour (1997), Staiger and Stock (1997) (SS), Stock and Wright (2000) (SW), Kleibergen (2002, 2005), Moreira (2003), and other papers referenced in Andrews and Stock (2007), is related to the present paper because it considers weak identification. The SS and SW papers are similar to the present paper because they analyze the behavior of estimators as well as tests and CS's. The SW and Kleibergen (2005) papers are similar because they consider nonlinear models, as does the present paper.

In the weak IV literature, the criterion functions considered are not indexed by the parameters that are the source of weak identification. Thus, in linear IV models, the reduced form parameters do not appear in the criterion function. Similarly, in SW, which applies to nonlinear models, high-level conditions are placed on the population moment functions under which the IV's are weak for some parameters. On the other hand, in the present paper, the potential source of weak identification is an explicit part of the model.[6] In consequence, the present paper and the weak IV literature are complementary—they focus on different criterion functions/models.

However, there is some overlap. For example, in the standard linear IV regression model, the criterion function for the limited information maximum likelihood (LIML) estimator can be written either as (i) a function of the parameters in the structural equation plus the parameters in the accompanying reduced-form equations, which fits the framework of the present paper, or (ii) a function of the structural equation parameters only via concentrating out the reduced-form parameters, as in the analysis in Anderson and Rubin (1949) and Staiger and Stock (1997).[7]

The focus of the present paper and many papers in the weak IV literature is somewhat different. We are concerned with cases in which the model is strongly identified in part of the parameter space, unidentified or weakly identified or semi-strongly identified in another part of the parameter space and the researcher does not know which case obtains. In contrast, the weak IV literature is focussed more on the weakly-identified case.

In consequence, we analyze the full range of strength-of-identification scenarios and provide methods that are suitable for sub-vectors and low dimensional functions, $r(\theta)$, of the full parameter vector θ under semi-strong and strong identification and are robust to weak identification. These methods allow for asymptotically efficient procedures when the identification is semi-strong or strong.

In contrast, most papers in the weak IV literature employ Anderson-Rubin-type procedures which yield inference concerning the whole parameter vector θ. To obtain inference for sub-vectors or functions $r(\theta)$ of θ, one uses some auxiliary method, such as projection or Bonferroni's inequality. This approach leads to asymptotically conservative procedures for sub-vectors or functions $r(\theta)$ in both weakly- and strongly-identified scenarios. Kleibergen and Mavroeidis (2009) analyze sub-vector methods in moment condition models with weak IV's and show that these methods have correct asymptotic

[6] To help clarify the differences, we show in Appendix E of AC1-SM that SW's Assumption C fails in the nonlinear regression model when a nonlinear regression parameter is weakly identified because its corresponding multiplicative coefficient is close to zero.

[7] The same is true of the two-stage least squares (2SLS) estimator. The 2SLS estimator fits the framework of the present paper by writing the criterion function for the structural and reduced-form parameters as a single GMM criterion function with no over-identifying restrictions.

null rejection probabilities under weak IV's, i.e., probabilities less than or equal to the nominal level α, but typically are asymptotically conservative.

The finite-sample results of Dufour (1997) and Gleser and Hwang (1987) for CS's and tests are applicable to the models considered in this paper. This paper considers the case where the potentially unidentified parameter π lies in a bounded set Π. In this case, Cor. 3.4 of Dufour (1997) implies that if the diameter of a CS for π is as large as the diameter of Π with probability less than $1 - 2\alpha$ then the CS has (exact) size less than $1 - \alpha$ (under certain assumptions).

Antoine and Renault (2009, 2010) consider GMM estimation with instruments that lie in what we call the semi-strong category. Their emphasis is on asymptotic efficiency with semi-strong instruments, rather than the behavior of statistics across the full range of strengths of identification as is considered here.

Nelson and Startz (2007) introduces the zero-information-limit condition, which applies to the models considered in this paper, and discuss its implications. Ma and Nelson (2006) considers tests based on linearization for models of the type considered in this paper. Neither of these papers establishes the large sample properties of estimators, tests, and CS's along the lines given in this paper.

Sargan (1983) provides asymptotic results for linear-in-variables and nonlinear-in-parameters simultaneous equations models in which some parameters are unidentified. Phillips (1989) and Choi and Phillips (1992) provide finite-sample and asymptotic results for linear simultaneous equations and linear spurious regression models in which some parameters are unidentified. Their results do not overlap very much with those in this paper because the present paper is focussed on nonlinear models. Their asymptotic results are pointwise in the parameters, which covers the unidentified- and strongly-identified categories, but not the weakly-identified and semi-strongly-identified categories described above.

The results of the present paper apply to the nonlinear regression model estimated by LS. We use this as an example to illustrate the general results of the paper, see AC1-SM. In the example, the regressors are i.i.d. or stationary and ergodic. One also can apply the approach of this paper to the case where the regressors are integrated. In this case, the general results given below do not apply directly. However, by using the asymptotics for nonlinear and nonstationary processes developed by Park and Phillips (1999, 2001), the approach goes through, as shown recently by Shi and Phillips (2009). With integrated regressors, the nonlinear regression model is a nonlinear cointegration model. Shi and Phillips (2009) employs the same method of computing asymptotic size and of constructing identification-robust CS's as was introduced in an early version of this paper and Cheng (2008).

The remainder of the paper is organized as follows. Section 2 describes the method used in the paper to obtain the asymptotic results. Section 3 introduces the extremum estimators, criterion functions, tests, confidence sets, and drifting sequences of distributions considered in the paper. Section 4 states the high-level assumptions employed. Section 5 provides the asymptotic results for the extremum estimators. Section 6 establishes the asymptotic distribution of t statistics and determines the asymptotic size of standard t CS's. Section 7 introduces methods of constructing robust tests and CS's whose asymptotic size equals to their nominal size and applies them to t tests and CI's. Section 8 introduces quasi-likelihood ratio (QLR) tests and CS's, which are considered in the numerical results for the ARMA(1, 1) example. For brevity, theoretical results for the QLR procedures are given in AC2. Section 9 provides asymptotic and finite-sample numerical results for the ARMA(1, 1) model. Appendix A of AC1-SM gives sufficient conditions for some of the high-level conditions stated in Section 4. Appendix B of AC1-SM provides proofs of the results given in Sections 5 and 6. Appendix C of AC1-SM verifies the assumptions of the paper for the ARMA example. Appendix D of AC1-SM provides additional Monte Carlo simulation results for the ARMA example. Appendix E of AC1-SM verifies the assumptions of the paper for the nonlinear regression example.

AC2 provides primitive sufficient conditions for the high-level assumptions of this paper for the class of estimators based on sample averages that are smooth functions of the parameter θ, which includes ML and LS estimators. It also provides general results for QLR tests and CS's for multi-dimensional hypotheses. AC3 provides sufficient conditions for the high-level assumptions for the class of GMM estimators and provides general results for Wald tests.

All limits below are taken "as $n \to \infty$." Let $o_{p\pi}(1)$, $O_{p\pi}(1)$, and $o_{\pi}(1)$ denote terms that are $o_p(1)$, $O_p(1)$, and $o(1)$, respectively, uniformly over a parameter $\pi \in \Pi$. Thus, $X_n(\pi) = o_{p\pi}(1)$ means that $\sup_{\pi \in \Pi} ||X_n(\pi)|| = o_p(1)$, where $||\cdot||$ denotes the Euclidean norm. Let "for all $\delta_n \to 0$" abbreviate "for all sequences of positive scalar constants $\{\delta_n : n \geq 1\}$ for which $\delta_n \to 0$." Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues, respectively, of a matrix A. All vectors are column vectors. For notational simplicity, we often write (a, b) instead of $(a', b')'$ for vectors a and b. Also, for a function $f(c)$ with $c = (a, b)$ $(= (a', b')')$, we often write $f(a, b)$ instead of $f(c)$. Let

0_d denote a d-vector of zeros. Because it arises frequently, we let 0 denote a d_β-vector of zeros, where d_β is the dimension of a parameter β. Let $R_{[\pm\infty]} = R \cup \{\pm\infty\}$. Let $R^p_{[\pm\infty]} = R_{[\pm\infty]} \times ... \times R_{[\pm\infty]}$ with p copies.

Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes indexed by $\pi \in \Pi$ for some space Π. The definition of weak convergence of R^v-valued functions on Π requires the specification of a metric d on the space $\mathcal{E}_v$ of R^v-valued functions on Π. We take d to be the uniform metric. The literature contains several definitions of weak convergence. We use any of the definitions that is compatible with the use of the uniform metric and for which the continuous mapping theorem (CMT) holds. These include the definitions employed by Pollard (1984, p. 65), Pollard (1990, p. 44), and van der Vaart and Wellner (1996, p. 17). The CMT's that correspond to these definitions are given by Pollard (1984, p. 70), Pollard (1990, p. 46), and van der Vaart and Wellner (1996, Thm. 1.3.6, p. 20). In the event of measurability issues, outer probabilities are used below implicitly in place of probabilities.

2. Description of Approach

The criterion functions/models considered in this paper possess the following characteristics:
(i) the criterion function does not depend on π when $\beta = 0$,
(ii) the criterion function viewed as a function of ψ with π fixed has a (stochastic) quadratic approximation wrt ψ (for ψ close to the true value of ψ) for each $\pi \in \Pi$ when the true β is close to the non-identification value 0 (see Assumption C1 in Section 4.4 below),
(iii) the (generalized) first derivative of this quadratic expansion converges weakly as a process indexed by $\pi \in \Pi$ to a Gaussian process after suitable normalization,
(iv) the (generalized) Hessian of this quadratic expansion is nonsingular asymptotically for all $\pi \in \Pi$ after suitable normalization,
(v) the criterion function viewed as a function of θ has a (stochastic) quadratic approximation wrt θ (for θ close to the true value) whether or not the true β is close to the non-identification value 0 (see Assumption D1 in Section 4.5 below),
(vi) the (generalized) first derivative of this quadratic expansion has an asymptotic normal distribution, where a matrix rescaling is employed when β is local to the non-identification value 0, and

(vii) the (generalized) Hessian of this quadratic expansion is nonsingular asymptotically, where a matrix rescaling is used when β is local to the non-identification value 0.

Now, we describe the approach used to establish the asymptotic results discussed in the Introduction. The estimator $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{\zeta}_n, \widehat{\pi}_n)$ is defined to minimize a criterion function $Q_n(\theta)$ over $\theta \in \Theta$. Let $\theta_n = (\beta_n, \zeta_n, \pi_n)$ denote the true parameter.

Several steps are employed. The first three steps apply to sequences of true parameters in Categories I and II.

Step 1. We consider the concentrated estimator $\widehat{\psi}_n(\pi)$ that minimizes $Q_n(\theta) = Q_n(\psi, \pi)$ over ψ for fixed $\pi \in \Pi$ and the concentrated criterion function $Q_n^c(\pi) = Q_n(\widehat{\psi}_n(\pi), \pi)$. We show that $\widehat{\psi}_n(\pi)$ is consistent for ψ_n uniformly over $\pi \in \Pi$. The method of proof is a variation of a standard consistency proof for extremum estimators adjusted to yield uniformity over π. The proof is analogous to that used in Andrews (1993) for estimators of structural change models in the situation where no structural change occurs.

Step 2. We employ a stochastic quadratic expansion of $Q_n(\psi, \pi)$ in ψ for given π about the non-identification point $\psi = \psi_{0,n} = (0, \zeta_n)$, rather than the true value ψ_n, which is key. By expanding about $\psi_{0,n}$, the leading term of the expansion, $Q_n(\psi_{0,n}, \pi)$, does not depend on π because $Q_n(\beta, \zeta, \pi)$ does not depend on π when $\beta = 0$. For each $\pi \in \Pi$, we obtain a linear approximation to $\widehat{\psi}_n(\pi)$ after centering around $\psi_{0,n}$ and rescaling. At the same time, we obtain a quadratic approximation of $Q_n^c(\pi)$. Both results hold uniformly in π. The method employed has two steps.

The first step of the two-step method involves establishing a rate of convergence result for $\widehat{\psi}_n(\pi) - \psi_{0,n}$. The second step uses this rate of convergence result to obtain the linear approximation of $\widehat{\psi}_n(\pi) - \psi_{0,n}$ (after rescaling) and the quadratic approximation of $Q_n(\psi, \pi) - Q_n(\psi_{0,n}, \pi)$ (after rescaling) as a function of ψ. Because $Q_n(\psi_{0,n}, \pi)$ does not depend on π, it does not effect the behavior of $\widehat{\psi}_n(\pi)$ or $\widehat{\pi}_n$. The two-step method used here is like that used by Chernoff (1954), Pakes and Pollard (1989), and Andrews (1999) among others, except that it is carried out for a family of values π, as in Andrews (2001), rather than a single value, and the results hold uniformly over π.

Step 3. We determine the asymptotic behavior of the (generalized) first derivative of $Q_n(\psi, \pi)$ wrt ψ evaluated at $\psi_{0,n}$. Due to the expansion about $\psi_{0,n}$, rather than about the true value ψ_n, a bias is introduced in the first derivative—its mean is not zero. The results here differ between Category I and II sequences. With Category I sequences, one obtains a stochastic term (a mean zero Gaussian process indexed by π) plus a non-stochastic term due to the bias ($K(\pi;\gamma_0)b$ in the notation used below) and the two are of the same order of magnitude. With Category II sequences, the true β_n is farther from the point of expansion 0 than with Category I sequences and, in consequence, the non-stochastic bias term is of a larger order of magnitude than the stochastic term. In this case, the limit is non-stochastic.

We also determine the asymptotic behavior of the (generalized) Hessian matrix of $Q_n(\psi,\pi)$ wrt ψ evaluated at $\psi_{0,n}$. It has a non-stochastic limit. There is no problem here with singularity of the Hessian because it is the Hessian for ψ only, not $\theta = (\psi,\pi)$, and ψ is identified.

For Category I sequences, the results of this step combined with those of Step 2 and the condition $n^{1/2}(\psi_n - \psi_{0,n}) \rightarrow (b,0)$ gives the asymptotic distribution of (i) the concentrated estimator $\widehat{\psi}_n(\cdot)$ viewed as a stochastic process indexed by $\pi \in \Pi$: $n^{1/2}(\widehat{\psi}_n(\cdot) - \psi_n) \Rightarrow \tau(\cdot)$, where $\tau(\cdot)$ is a Gaussian process indexed by $\pi \in \Pi$ whose mean is non-zero unless $b = 0$, and (ii) the concentrated criterion function $Q_n^c(\cdot)$: $n(Q_n^c(\cdot) - Q_n(\psi_{0,n},\pi)) \Rightarrow \xi(\cdot)$, where $\xi(\cdot)$ is a quadratic form in $\tau(\cdot)$.

For Category II sequences, putting the results above together yields: (i) a rate of convergence result for $\widehat{\psi}_n(\pi)$: $\sup_{\pi\in\Pi} ||\widehat{\psi}_n(\pi) - \psi_{0,n}|| = O_p(||\beta_n||)$ that is just fast enough to obtain a rate of convergence result for $\widehat{\psi}_n - \psi_n$ in Step 6 below and (ii) the (non-stochastic) probability limit $\eta(\pi)$ of $Q_n^c(\pi)$ (after normalization): $||\beta_n||^{-1}(Q_n^c(\pi) - Q_n(\psi_{0,n},\pi)) \rightarrow_p \eta(\pi)$ uniformly over $\pi \in \Pi$.

Step 4. For Category I sequences, we use $\widehat{\pi}_n = \arg\min_{\pi\in\Pi} Q_n^c(\pi)$, $n(Q_n^c(\cdot) - Q_n(\psi_{0,n},\pi)) \Rightarrow \xi(\cdot)$ from Step 3 (where $Q_n(\psi_{0,n},\pi)$ does not depend on π), and the continuous mapping theorem (CMT) to obtain $\widehat{\pi}_n \rightarrow_d \pi^* = \arg\min_{\pi\in\Pi} \xi(\pi)$. In this case, $\widehat{\pi}_n$ is not consistent. Given the asymptotic distribution of $\widehat{\pi}_n$, the result $n^{1/2}(\widehat{\psi}_n(\cdot) - \psi_n) \Rightarrow \tau(\cdot)$ from Step 3, and the CMT, we obtain the asymptotic distribution of $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$: $n^{1/2}(\widehat{\psi}_n - \psi_n) \rightarrow_d \tau(\pi^*)$. This completes the asymptotic results for $(\widehat{\psi}_n, \widehat{\pi}_n)$ for Category I sequences of true parameters.

Step 5. For Category II sequences, we obtain the consistency of $\widehat{\pi}_n$ by using the uniform convergence in probability of $Q_n^c(\pi)$ (after normalization) to the non-stochastic quadratic form, $\eta(\pi)$, established in Step 3, combined with the property that $\eta(\pi)$ is uniquely minimized at the limit π_0 of the true values π_n. The vector that appears in the quadratic form $\eta(\pi)$ is the vector of biases of the (generalized) first derivative obtained in Step 3, which appears due to the expansion around $\psi_{0,n}$ rather than around ψ_n. The weight matrix of $\eta(\pi)$ is the inverse of the Hessian discussed in Step 3.

Step 6. For Category II sequences, we use the rate of convergence result $\sup_{\pi\in\Pi} ||\widehat{\psi}_n(\pi) - \psi_{0,n}|| = O_p(||\beta_n||)$ from Step 3 and a relationship between the bias of the (generalized) first-derivative and the (generalized) Hessian (wrt ψ) to obtain a rate of convergence result for $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$ centered at the true value ψ_n: $\widehat{\psi}_n - \psi_n = o_p(||\beta_n||)$.

Step 7. For Category II and III sequences, we carry out stochastic quadratic expansions of $Q_n(\theta)$ about the true value θ_n. The argument proceeds as in Step 2 (but the expansion here is in θ, not in ψ with π fixed, and the expansion is about the true value). First, we obtain a rate of convergence result for $\widehat{\theta}_n - \theta_n$ and then with this rate we obtain the asymptotic distribution of $\widehat{\theta}_n - \theta_n$ (after rescaling) using the quadratic approximation of $Q_n(\theta)$ in a particular neighborhood of θ_n. The result obtained is consistency and asymptotic normality (with mean zero) for $\widehat{\theta}_n$ with rate $n^{1/2}$ for $\widehat{\psi}_n$ for Category II and III sequences, rate $n^{1/2}$ for $\widehat{\pi}_n$ for Category III sequences, and rate $n^{1/2}||\beta_n||$ ($<< n^{1/2}$) for $\widehat{\pi}_n$ for Category II sequences. The last rate result is due to the convergence of β_n to 0 albeit slowly. With Category II sequences, $\widehat{\pi}_n$ is consistent and asymptotically normal but with a slower rate of convergence than is standard.

For Category II sequences, the results in this step are complicated by two issues. First, the (generalized) Hessian matrix for θ with the standard normalization is singular asymptotically because $\beta_n \to 0$ and the random criterion function $Q_n(\theta)$ becomes more flat wrt π for β in a neighborhood of β_n the closer is β_n to 0. This requires a matrix rescaling of the Hessian based on the magnitude of $||\beta_n||$. Second, the quadratic approximation of the criterion function wrt θ around the true value θ_n only holds for θ close enough to θ_n; specifically, only for $\theta \in \Theta_n(\delta_n) = \{\theta \in \Theta : ||\psi - \psi_n|| \leq \delta_n ||\beta_n|| \,\&\, ||\pi - \pi_n|| \leq \delta_n\}$ for constants $\delta_n \to 0$. Thus, ψ needs to be very close to the true value ψ_n for the quadratic approximation to hold. It is for this reason that the rate of convergence result $\widehat{\psi}_n - \psi_n = o_p(||\beta_n||)$ in Step 6 is a key result. The quadratic approximation requires $\theta \in \Theta_n(\delta_n)$ because for such $\theta = (\beta, \zeta, \pi)$ we have $||\beta||/||\beta_n|| = 1 + o(1)$ and, hence, the rescaling that enters the Hessian is asymptotically equivalent whether it is based on β or the true value β_n. (For example, see the verification of Assumption Q1(iv) for the LS example in (15.17) to see that the restriction $\theta \in \Theta_n(\delta_n)$ is required for the quadratic approximation to hold in this example.)

Step 8. We obtain the asymptotic null distributions of t test statistics for linear and nonlinear restrictions using the asymptotic distributions of the estimators described in

Steps 1-7 plus asymptotic results for the variance matrix and standard error estimators upon which the test statistics depend. The latter exhibit non-standard behavior for Category I sequences because $\widehat{\pi}_n$ is random even in the limit. These results yield the asymptotic null rejection probabilities and coverage probabilities of standard t test for Category I-III sequences.

For Category I sequences, the asymptotic distribution of the t statistic for a linear or nonlinear restriction that involves both π and ψ is found to depend only on the randomness in $\widehat{\pi}_n$ and not on the randomness in $\widehat{\psi}_n$. This occurs because the former is of a larger order of magnitude than the latter. When a restriction does not involve π, then the asymptotic null distribution of the t statistic for Category I sequences usually still depends on the (asymptotically non-standard) randomness of $\widehat{\pi}_n$ through the standard deviation estimator and implicitly through the effect of the randomness of $\widehat{\pi}_n$ on the asymptotic distribution of $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$.

Step 9. Using the asymptotic results from Step 8 for Category I-III sequences of true parameters, combined with the argument from Andrews and Guggenberger (2010), as formulated in Andrews, Cheng, and Guggenberger (2009), we obtain a formula for the asymptotic size of standard t tests and CS's. Their behavior under Category I sequences determines whether a test over-rejects asymptotically and whether a CS under-covers asymptotically. Under Category II and III sequences, they perform asymptotically as desired.

Step 10. We introduce LF and data-dependent robust critical values that yield tests and CI's that have correct asymptotic size even in the presence of identification failure and weak identification in part of the parameter space. The adjusted critical values employ the asymptotic formulae derived in Steps 8 and 9.[8]

3. Estimator and Criterion Function

3.1. Extremum Estimators

We consider an estimator $\widehat{\theta}_n$, such as an ML, LS, quantile, GMM, GEL, or MD estimator, that is defined by minimizing a sample criterion function. The sample criterion

[8]Steps 1-9 correspond to the following results: Step 1, Lemma 5.1; Step 2, Lemma 12.2; Step 3, Lemmas 12.1 and 12.2; Step 4, Theorem 5.1; Step 5, Lemma 5.3; Step 6, Lemmas 12.3 and 12.4; Step 7, Theorem 5.2; Step 8, Theorem 6.1; Step 9, Theorem 6.2; and Step 10, Theorem 7.1.

function, $Q_n(\theta)$, depends on the observations $\{W_i : i \leq n\}$, which may be i.i.d., i.n.i.d., or temporally dependent.[9]

The paper focuses on inference when θ is not identified (by the criterion function $Q_n(\theta)$) at some points in the parameter space. Lack of identification occurs when the $Q_n(\theta)$ is flat wrt some sub-vector of θ. To model this identification problem, θ is partitioned into three sub-vectors:

$$\theta = (\beta, \zeta, \pi) = (\psi, \pi), \text{ where } \psi = (\beta, \zeta). \tag{3.1}$$

The parameter $\pi \in R^{d_\pi}$ is unidentified when $\beta = 0$ ($\in R^{d_\beta}$). The parameter $\psi = (\beta, \zeta) \in R^{d_\psi}$ is always identified. The parameter $\zeta \in R^{d_\zeta}$ does not effect the identification of π. These conditions are stated more precisely in Assumptions A and B3 below. They allow for a wide range of cases, including cases in which reparametrization is used to convert a model into the framework considered here.

Example 1. We consider an ARMA(1, 1) model. We use it as a running example to illustrate the more general results. In this model, the AR and MA parameters are not identified when their values are equal. This occurs when the ARMA(1, 1) time series is serially uncorrelated—a case of considerable interest in many practical applications. Simulation results in Ansley and Newbold (1980) and Nelson and Startz (2007) demonstrate that this causes substantial bias, variance, and size problems when the AR and MA parameters are close in value. We provide a comprehensive asymptotic analysis of the problem.

The observed ARMA(1, 1) time series $\{Y_t : 0 \leq t \leq n\}$ is generated by the following equation:

$$Y_t = (\pi_0 + \beta_0) Y_{t-1} + \varepsilon_t - \pi_0 \varepsilon_{t-1} \text{ for } t = ..., 0, 1, ..., \tag{3.2}$$

where the innovations $\{\varepsilon_t : t = ..., 0, 1, ..\}$ are i.i.d. with mean zero and variance ζ_0, and the distribution of $\zeta_0^{-1/2} \varepsilon_t$ is ϕ_0. The true MA parameter is π_0 and the true AR parameter is $\pi_0 + \beta_0$. For notational simplicity, we sometimes write $\rho_0 = \pi_0 + \beta_0$ and $\rho = \pi + \beta$. When $\beta_0 = 0$, the model is $Y_t = \pi_0 Y_{t-1} + \varepsilon_t - \pi_0 \varepsilon_{t-1}$, which is equivalent to $Y_t = \varepsilon_t$. In this case, ρ_0 and π_0 are not identified.

We consider the Gaussian quasi-log likelihood function for $\theta = (\beta, \zeta, \pi)$ conditional

[9] The indices i and t are inter-changeable in this paper. For the general results and cross-section examples, the observations are indexed by i $(= 1, ..., n)$. To conform with standard notation, the observations are indexed by t $(= 1, ..., n$ or $= -r, ..., n$ for some $r \geq 0)$ in time series examples, such as the ARMA(1, 1) example.

on Y_0 and ε_0. The conditioning value ε_0 is asymptotically negligible, so for simplicity (and wlog for the asymptotic results) we set $\varepsilon_0 = Y_0$ in the log likelihood. The (conditional) QML criterion function for $\theta = (\beta, \zeta, \pi)'$ (multiplied by $-n^{-1}$ and ignoring a constant) is

$$Q_n(\theta) = \frac{1}{2}\log\zeta + \frac{1}{2\zeta}n^{-1}\sum_{t=1}^{n}\left(Y_t - \beta\sum_{j=0}^{t-1}\pi^j Y_{t-j-1}\right)^2. \tag{3.3}$$

(See Appendix C of AC1-SM for details regarding its calculation.) The criterion function $Q_n(\theta)$ does not depend on π when $\beta = 0$.

The results for this example can be extended to the case where the mean of the strictly stationary time series Y_t is μ_0. In this case, (3.2) holds with Y_t and Y_{t-1} replaced by $Y_t - \mu_0$ and $Y_{t-1} - \mu_0$, respectively. The mean μ_0 can be estimated by ML, in which case Y_t is replaced by $Y_t - \mu$ in the criterion function and the criterion function is minimized wrt μ as well as the other parameters, or μ_0 can be estimated by $\overline{Y}_n = n^{-1}\sum_{t=1}^{n} Y_t$, in which case Y_t is replaced by $Y_t - \overline{Y}_n$ in the criterion function. In either case, the asymptotic results concerning (β, ζ, π) are the same whether or not μ_0 is estimated, due to the block diagonality of the information matrix between μ and (β, ζ, π). □

The true distribution of the observations $\{W_i : i \leq n\}$ is denoted F_γ for some parameter $\gamma \in \Gamma$. We let P_γ and E_γ denote probability and expectation under F_γ. The parameter space Γ for the true parameter, referred to as the "true parameter space," is compact and is of the form:

$$\Gamma = \{\gamma = (\theta, \phi) : \theta \in \Theta^*, \phi \in \Phi^*(\theta)\}, \tag{3.4}$$

where the true parameter space for θ, Θ^*, is a compact subset of R^{d_θ} and $\Phi^*(\theta) \subset \Phi^*$ $\forall\theta \in \Theta^*$ for some compact metric space Φ^* with a metric that induces weak convergence of the bivariate distributions (W_i, W_{i+m}) for all $i, m \geq 1$.[10,11] In unconditional likelihood scenarios, no parameter ϕ appears. In conditional likelihood scenarios, with conditioning variables $\{X_i : i \geq 1\}$, ϕ indexes the distribution of $\{X_i : i \geq 1\}$. In moment condition models, θ is a finite-dimensional parameter that appears in the moment functions and

[10] That is, the metric satisfies: if $\gamma \to \gamma_0$, then (W_i, W_{i+m}) under γ converges in distribution to (W_i, W_{i+m}) under γ_0. Note that Γ is a metric space with metric $d_\Gamma(\gamma_1, \gamma_2) = ||\theta_1 - \theta_2|| + d_{\Phi^*}(\phi_1, \phi_2)$, where $\gamma_j = (\theta_j, \phi_j) \in \Gamma$ for $j = 1, 2$ and d_{Φ^*} is the metric on Φ^*.

[11] The asymptotic results below give uniformity results over the parameter space Γ. If one is interested in a non-compact parameter space Φ_1^* for the parameter ϕ, instead of Φ^*, then one can apply the results established here to show that the uniformity results hold for all compact subsets Φ^* of Φ_1^* that satisfy the given conditions.

ϕ indexes those aspects of the distribution of the observations that are not determined by θ. In nonlinear regression models estimated by least squares, θ indexes the regression functions and possibly a finite-dimensional feature of the distribution of the errors, such as its variance, and ϕ indexes the remaining characteristics of the distribution of the errors, which may be infinite dimensional.

By definition, the extremum estimator $\widehat{\theta}_n$ (approximately) minimizes $Q_n(\theta)$ over an "optimization parameter space" Θ:[12]

$$\widehat{\theta}_n \in \Theta \text{ and } Q_n(\widehat{\theta}_n) = \inf_{\theta \in \Theta} Q_n(\theta) + o(n^{-1}). \tag{3.5}$$

We assume that the interior of the optimization parameter space Θ includes the true parameter space Θ^* (see Assumption B1 below). This ensures that the asymptotic distribution of $\widehat{\theta}_n$ is not affected by boundary constraints for any sequence of true parameters in Θ^*. The focus of this paper is not on the effects of boundary constraints.

3.2. Confidence Sets and Tests

We are interested in the effect of lack of identification or weak identification on the behavior of the extremum estimator $\widehat{\theta}_n$. In addition, we are interested in its effects on CS's for various functions $r(\theta)$ of θ and on tests of null hypotheses of the form $H_0 : r(\theta) = v$.

A CS is obtained by inverting a test. For example, a nominal $1 - \alpha$ CS for $r(\theta)$ is

$$CS_n = \{v : \mathcal{T}_n(v) \leq c_{n,1-\alpha}(v)\}, \tag{3.6}$$

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

[12] The $o(n^{-1})$ term in (3.5), and in (5.1) and (5.2) below, is a fixed sequence of constants that does not depend on the true parameter $\gamma \in \Gamma$ and does not depend on π in (5.1). The $o(n^{-1})$ term makes it clear that the infima in these equations need not be achieved exactly. This allows for some numerical inaccuracy in practice and also circumvents the issue of the existence of parameter values that achieve the infima. In contrast to many results in the extremum estimator literature, the $o(n^{-1})$ term is not a random $o_p(n^{-1})$ term here because a quantity is $o_p(n^{-1})$ only for a specific sequence of true distributions and the uniform results given below require properties of the extremum estimators to hold for arbitrary sequences of true distributions.

of a CS for $r(\theta)$ is

$$P_\gamma(r(\theta) \in CS_n) = P_\gamma(\mathcal{T}_n(r(\theta)) \leq c_{n,1-\alpha}(r(\theta))), \tag{3.7}$$

where $P_\gamma(\cdot)$ denotes probability when γ is the true value.

The paper focuses on the smallest finite-sample coverage probability of a CS over the parameter space, i.e., the finite-sample size of the CS. It is approximated by the asymptotic size, which is defined to be

$$AsySz = \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} P_\gamma(r(\theta) \in CS_n) = \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} P_\gamma(\mathcal{T}_n(r(\theta)) \leq c_{n,1-\alpha}(r(\theta))). \tag{3.8}$$

For a test, we are interested in its null rejection probabilities and in particular its maximum null rejection probability, which is the size of the test. A test's asymptotic size is an approximation to the latter. The null rejection probabilities and asymptotic size of a test are given by

$$\begin{aligned} &P_\gamma(\mathcal{T}_n(v) > c_{n,1-\alpha}(v)) \text{ for } \gamma = (\theta, \phi) \in \Gamma \text{ with } r(\theta) = v \text{ and} \\ &AsySz = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma: r(\theta)=v} P_\gamma(\mathcal{T}_n(v) > c_{n,1-\alpha}(v)). \end{aligned} \tag{3.9}$$

3.3. Drifting Sequences of Distributions

In (3.8) and (3.9), the uniformity over $\gamma \in \Gamma$ for any given sample size n is crucial for the asymptotic size to be a good approximation to the finite-sample size. The value of γ at which the finite-sample size of a CS or test is attained may vary with the sample size. Therefore, to determine the asymptotic size we need to derive the asymptotic distribution of the test statistic $\mathcal{T}_n(v_n)$ under sequences of true parameters $\gamma_n = (\theta_n, \phi_n)$ and $v_n = r(\theta_n)$ that may depend on n.

Similarly, to investigate the finite-sample behavior of the extremum estimator under weak identification, we need to consider its asymptotic behavior under drifting sequences of true distributions—as in Staiger and Stock (1997), Stock and Wright (2000), and numerous other papers that consider weak instruments.

Results in Andrews and Guggenberger (2010) and Andrews, Cheng, and Guggenberger (2009) show that the asymptotic size of CS's and tests are determined by certain drifting sequences of distributions. In this paper, the following sequences $\{\gamma_n\}$ are key:

$$\begin{aligned}
\Gamma(\gamma_0) &= \{\{\gamma_n \in \Gamma : n \geq 1\} : \gamma_n \to \gamma_0 \in \Gamma\}, \qquad (3.10)\\
\Gamma(\gamma_0, 0, b) &= \left\{\{\gamma_n\} \in \Gamma(\gamma_0) : \beta_0 = 0 \text{ and } n^{1/2}\beta_n \to b \in R^{d_\beta}_{[\pm\infty]}\right\}, \text{ and}\\
\Gamma(\gamma_0, \infty, \omega_0) &= \left\{\{\gamma_n\} \in \Gamma(\gamma_0) : n^{1/2}||\beta_n|| \to \infty \text{ and } \beta_n/||\beta_n|| \to \omega_0 \in R^{d_\beta}\right\},
\end{aligned}$$

where $\gamma_0 = (\beta_0, \zeta_0, \pi_0, \phi_0)$ and $\gamma_n = (\beta_n, \zeta_n, \pi_n, \phi_n)$. Note that the 0 in $\Gamma(\gamma_0, 0, b)$ and the ∞ in $\Gamma(\gamma_0, \infty, \omega_0)$ stand for different things. In the former, $\beta_0 = 0$, and in the latter $n^{1/2}||\beta_n|| \to \infty$.

The sequences in $\Gamma(\gamma_0, 0, b)$ are in Categories I and II and are sequences for which $\{\beta_n\}$ is *close* to 0: $\beta_n \to 0$. When $||b|| < \infty$, $\{\beta_n\}$ is within $O(n^{-1/2})$ of 0 and the sequence is in Category I. The sequences in $\Gamma(\gamma_0, \infty, \omega_0)$ are in Categories II and III and are more *distant* from $\beta = 0$: $n^{1/2}||\beta_n|| \to \infty$. The sets $\Gamma(\gamma_0, 0, b)$ and $\Gamma(\gamma_0, \infty, \omega_0)$ are *not* disjoint. Both contain sequences in Category II.

Throughout the paper we use the terminology: "under $\{\gamma_n\} \in \Gamma(\gamma_0)$" to mean "when the true parameters are $\{\gamma_n\} \in \Gamma(\gamma_0)$ for any $\gamma_0 \in \Gamma$;" "under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$" to mean "when the true parameters are $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for any $\gamma_0 \in \Gamma$ with $\beta_0 = 0$ and any $b \in R^{d_\beta}_{[\pm\infty]}$;" and "under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$" to mean "when the true parameters are $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ for any $\gamma_0 \in \Gamma$ and any $\omega_0 \in R^{d_\beta}$ with $||\omega_0|| = 1$."

4. Assumptions

This section provides the high-level conditions under which the results of the paper hold. Verification of the high-level conditions is illustrated using the running example of ML estimation of the ARMA(1, 1) model. Appendix E of AC1-SM verifies the high-level conditions in a cross-section nonlinear regression model. Furthermore, various sets of primitive sufficient conditions for the high-level conditions are given for different types of estimators in Appendix A of AC1-SM, AC2, and AC3. AC2 considers sample average criterion functions, such as ML and LS, that are smooth functions of θ. AC3 considers GMM and MD criterion functions.

4.1. Basic Identification Assumption

The first assumption specifies that θ is not identified (via the criterion function $Q_n(\theta)$) at some points in the parameter space.

Assumption A. If $\beta = 0$, $Q_n(\theta)$ does not depend on π, $\forall \theta = (\beta, \zeta, \pi) \in \Theta$, $\forall n \geq 1$, for any true parameter $\gamma^* \in \Gamma$.[13]

Assumption A specifies that $Q_n(\theta)$ is flat in π when $\beta = 0$. This flatness causes identification failure for π when $\beta = 0$ because $Q_n(\theta)$ cannot distinguish $\theta = (0, \zeta, \pi^*)$ from $\theta' = (0, \zeta, \pi)$ for any $\pi \in \Pi$. The non-identification of π invalidates the standard consistency argument for an extremum estimator based on $Q_n(\theta)$ and causes non-standard asymptotic distributions of extremum estimators and corresponding test statistics. The situation considered in this paper belongs to a broad category of cases where test statistics have discontinuous asymptotic distributions wrt the true parameter value. Here the discontinuity happens when the true value β^* equals 0. It is worth mentioning that the flatness specified in Assumption A does not affect identification of π when the true value $\beta^* \neq 0$ because the extremum estimator $\widehat{\beta}_n$ of β^* is consistent and hence the minimum of the criterion function occurs at values of β where flatness in π does not occur.

Example 1 (cont.). For $Q_n(\theta)$ defined as in (3.3), Assumption A obviously holds. $\square$

4.2. Parameter Space Assumptions

Next, we specify conditions on the parameter spaces Θ and Γ. To obtain asymptotic size results for tests and CS's, the parameter space must be specified precisely. Without loss of generality (wlog), the optimization parameter space Θ can be written as

$$\begin{aligned} \Theta &= \{\theta = (\psi, \pi) : \psi \in \Psi(\pi), \pi \in \Pi\}, \text{ where} \\ \Pi &= \{\pi : (\psi, \pi) \in \Theta \text{ for some } \psi\} \text{ and} \\ \Psi(\pi) &= \{\psi : (\psi, \pi) \in \Theta\} \text{ for } \pi \in \Pi. \end{aligned} \tag{4.1}$$

We allow $\Psi(\pi)$ to depend on π and, hence, Θ need not be a product space between ψ and π. This is needed in the ARMA(1, 1) example among others.[14]

[13] Assumption A requires the stated condition to hold for all possible realizations of $Q_n(\cdot)$ for any true parameter $\gamma^* \in \Gamma$. Assumption A can be weakened to an a.s. requirement for each $\gamma^* \in \Gamma$, but there seems to be no gain in terms of applications of interest by doing so.

[14] We write Θ in terms of the sets Π and $\Psi(\pi)$, rather than sets Ψ and $\Pi(\psi)$, because below we carry out quadratic expansions of $Q_n(\psi, \pi)$ wrt ψ for each $\pi \in \Pi$ and this yields stochastic processes that are

Define $\Theta_\delta^* = \{\theta \in \Theta^* : ||\beta|| < \delta\}$, where Θ^* is the true parameter space for θ. The optimization parameter space Θ satisfies:

Assumption B1. (i) $int(\Theta) \supset \Theta^*$.
(ii) For some $\delta > 0$, $\Theta \supset \{\beta \in R^{d_\beta} : ||\beta|| < \delta\} \times \mathcal{Z}^0 \times \Pi \supset \Theta_\delta^*$ for some non-empty open set $\mathcal{Z}^0 \subset R^{d_\zeta}$ and Π as in (4.1).
(iii) Π is compact.

Because the optimization parameter space is user selected, Assumption B1 can be made to hold by the choice of Θ.[15] Assumption B1(ii) ensures that Θ is compatible with (i) a stochastic quadratic approximation of $Q_n(\theta) = Q_n(\psi, \pi)$ wrt ψ around $\psi^* = (0, \zeta^*)$ for each $\pi \in \Pi$, see Assumption C1 below, (ii) the empirical process $\{G_n(\pi) : \pi \in \Pi\}$ defined in Assumption C3 below, and (iii) the definition of $K_n(\theta; \gamma^*)$ in Assumption C5 below.

The true parameter space Γ satisfies:

Assumption B2. (i) Γ is compact and (3.4) holds.
(ii) $\forall \delta > 0$, $\exists \gamma = (\beta, \zeta, \pi, \phi) \in \Gamma$ with $0 < ||\beta|| < \delta$.
(iii) $\forall \gamma = (\beta, \zeta, \pi, \phi) \in \Gamma$ with $0 < ||\beta|| < \delta$ for some $\delta > 0$, $\gamma_a = (a\beta, \zeta, \pi, \phi) \in \Gamma$ $\forall a \in [0, 1]$.

Assumption B2(ii) guarantees that Γ is not empty and that there are elements γ of Γ whose β values are non-zero but are arbitrarily close to 0, which is the region of the true parameter space where near lack of identification occurs. Assumption B2(iii) ensures that Γ is compatible with the existence of partial derivatives of certain expectations wrt the true parameter β^* around $\beta^* = 0$. These partial derivatives arise in (4.16) and Assumption C5 below.

Example 1 (cont.). In the ARMA example, the optimization parameter space Θ is

$$\Theta = \{\theta = (\beta, \zeta, \pi)' : \beta \in [\rho_L - \pi, \rho_U - \pi],\ \zeta \in [\zeta_L, \zeta_U],\ \pi \in \Pi = [\pi_L, \pi_U]\}, \quad (4.2)$$

where $-1 < \rho_L < \pi_L < \pi_U < \rho_U < 1$ and $0 < \zeta_L < \zeta_U < \infty$. By the definition of Θ, the

indexed by the fixed set Π and that converge weakly as processes on Π.

[15] Assumption B1(iii) is used to show that certain continuous functions on Π introduced in Assumptions C6 and C7 below, which have unique minima on Π, satisfy "identifiable uniqueness" properties. Assumption B1(iii) could be avoided by imposing "identifiable uniqueness" properties directly in Assumptions C6 and C7.

autoregressive parameter $\rho = \pi + \beta$ lies in $[\rho_L, \rho_U]$.[16]

The true parameter space for θ is

$$\Theta^* = \{\theta = (\beta, \zeta, \pi)' : \beta \in [\rho_L^* - \pi, \rho_U^* - \pi],\ \zeta \in [\zeta_L^*, \zeta_U^*],\ \pi \in [\pi_L^*, \pi_U^*]\}, \tag{4.3}$$

where $\pi_L < \pi_L^* < \pi_U^* < \pi_U$, $\rho_L < \rho_L^* < \pi_L^* < \pi_U^* < \rho_U^* < \rho_U$, and $\zeta_L < \zeta_U^* < \zeta_U^* < \zeta_U$.

Let ξ_t denote the normalized innovation $\zeta^{-1/2}\varepsilon_t$, which has mean zero and variance one. The true parameter space for $\gamma = (\theta, \phi)$ is

$$\begin{aligned} &\Gamma = \{\gamma = (\theta, \phi) : \theta \in \Theta^*, \phi \in \Phi^*\},\ \text{where} \\ &\Phi^* \text{ is some compact subset of } \Phi \text{ wrt the metric } d_\Phi,\ \text{and} \\ &\Phi = \{\phi : E_\phi \xi_t = 0,\ E_\phi \xi_t^2 = 1,\ E_\phi(\xi_t^2 - 1)^2 \geq \delta_1,\ E_\phi|\xi_t|^{4+\delta_2} \leq K\} \end{aligned} \tag{4.4}$$

for some constants $\delta_1, \delta_2 > 0$ and $0 < K < \infty$, where d_Φ is some metric on the space of distributions on R that induces weak convergence.

With these definitions of Θ, Θ^*, and Γ, Assumptions B1 and B2 hold. $\square$

4.3. Criterion Function Limit Assumption

Here we specify the limit of the sample criterion function $Q_n(\theta)$ along drifting sequences of true parameters $\{\gamma_n\} \in \Gamma(\gamma_0)$ whose limit is $\gamma_0 \in \Gamma$.

Assumption B3. (i) For some non-stochastic real-valued function $Q(\theta; \gamma_0)$ on $\Theta \times \Gamma$,

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta; \gamma_0)| \to_p 0$$

under $\{\gamma_n\} \in \Gamma(\gamma_0)$, $\forall \gamma_0 \in \Gamma$.

(ii) When $\beta_0 = 0$, for every neighborhood Ψ_0 $(\subset R^{d_\psi})$ of $\psi_0 = (\beta_0, \zeta_0)$,

$$\inf_{\pi \in \Pi} \left(\inf_{\psi \in \Psi(\pi)/\Psi_0} Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0) \right) > 0,\ \forall \gamma_0 = (\psi_0, \pi_0, \phi_0) \in \Gamma.$$

[16] The conditions $\rho_L < \pi_L$ and $\pi_U < \rho_U$ imply that β can take values in a neighborhood of zero for any value of $\pi \in \Pi$.

(iii) When $\beta_0 \neq 0$, for every neighborhood Θ_0 $(\subset \Theta)$ of $\theta_0 = (\beta_0, \zeta_0, \pi_0)$,

$$\inf_{\theta \in \Theta / \Theta_0} Q(\theta; \gamma_0) - Q(\theta_0; \gamma_0) > 0, \ \forall \gamma_0 = (\theta_0, \phi_0) \in \Gamma.$$

Assumption B3(i) defines the (asymptotic) population criterion function $Q(\theta; \gamma_0)$. Assumption B3(ii) provides a condition for the identification of β and ζ despite the non-identification of π when $\beta_0 = 0$. Uniformity over Π is required due to the non-identification of π. A condition of this type also is used in Andrews (1993) for the uniform consistency of a family of estimators. A necessary condition for Assumption B3(ii) is that for any given $\pi \in \Pi$ and $\gamma_0 \in \Gamma$ with $\beta_0 = 0$, $Q(\psi, \pi; \gamma_0)$ is uniquely minimized by ψ_0. Assumption B3(iii) is a standard identification condition for θ when $\beta_0 \neq 0$. A condition of this sort is verified for various extremum estimators in Newey and McFadden (1994).

A set of primitive sufficient conditions for Assumptions B3(ii) and B3(iii) is given in Assumption B3* in Appendix A of AC1-SM.

Example 1 (cont.). In this example, the function $Q(\theta; \gamma_0)$ in Assumption B3(i) is

$$\begin{aligned} Q(\theta; \gamma_0) &= E_{\gamma_0} \rho_t(\theta), \text{ where} \\ \rho_t(\theta) &= \frac{1}{2} \log \zeta + \frac{1}{2\zeta} \left(Y_t - \beta \sum_{j=0}^{\infty} \pi^j Y_{t-j-1} \right)^2. \end{aligned} \tag{4.5}$$

The uniform convergence in Assumption B3(i) is established by showing pointwise convergence in probability via mean square convergence, stochastic equicontinuity, and boundedness of Θ.[17] For brevity, the details are given in AC1-SM. Assumptions B3(ii) and B3(iii) are verified by verifying the sufficient condition Assumption B3* given in Appendix A of AC1-SM. Again, for brevity, the details are given in Appendix C of AC1-SM. □

4.4. Close to $\beta = 0$ Assumptions

The following Assumptions C1-C8 are used to determine the asymptotic distributions of estimators and test statistics under sequences of true parameters $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$

[17] The sum over j in (4.5) runs to ∞, whereas that in (3.3) runs to $t-1$, because the initiation of the time series at $t = 0$ is asymptotically negligible.

with $||b|| < \infty$ and to establish the consistency of $\widehat{\pi}_n$ under sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| = \infty$. The "C" denotes that the sequences of parameters $\{\gamma_n\}$ considered are *close* to the point of non-identification.

The first assumption, Assumption C1, requires that the criterion function $Q_n(\theta)$ has a stochastic quadratic expansion in ψ around the non-identification point $\psi_{0,n} = (0, \zeta_n)$ uniformly in $\pi \in \Pi$. Assumptions C2, C3, and C8 concern the behavior of the (generalized) first derivative in the expansion. Assumption C4 concerns the behavior of the (generalized) second derivative. Assumptions C5 and C7 arise because the quadratic expansion is about the non-identification point $\psi_{0,n}$, rather than the true value ψ_n. Assumptions C6 and C7 are used when determining the asymptotic behavior of $\widehat{\pi}_n$.

We now define a sequence of scalar constants $\{a_n(\gamma_n) : n \geq 1\}$ that provides the normalization required so that the (generalized) first derivative in the quadratic expansion in Assumption C1 is non-degenerate asymptotically (see Lemma 12.1 in Appendix B of AC1-SM). These constants appear in the conditions on the remainder term of the approximation in Assumption C1. Define

$$a_n(\gamma_n) = \begin{cases} n^{1/2} & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } ||b|| < \infty \\ ||\beta_n||^{-1} & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } ||b|| = \infty. \end{cases} \tag{4.6}$$

Note that $||\beta_n||^{-1} < n^{1/2}$ for n large when $||b|| = \infty$, because $n^{1/2}||\beta_n|| \to \infty$.[18] Hence, $a_n(\gamma_n) \leq n^{1/2}$ for n large.

Assumption C1. Under $\{\gamma_n = (\beta_n, \zeta_n, \pi_n, \phi_n)\} \in \Gamma(\gamma_0, 0, b)$, for some $\delta > 0$, $\forall \theta = (\psi, \pi) \in \Theta_\delta = \{\theta \in \Theta : ||\beta|| < \delta\}$,
(i) the sample criterion function $Q_n(\psi, \pi)$ has a quadratic expansion in ψ around $\psi_{0,n} = (0, \zeta_n)$ for given π:

$$\begin{aligned} Q_n(\psi, \pi) = {} & Q_n(\psi_{0,n}, \pi) + D_\psi Q_n(\psi_{0,n}, \pi)'(\psi - \psi_{0,n}) + \\ & \frac{1}{2}(\psi - \psi_{0,n})' D_{\psi\psi} Q_n(\psi_{0,n}, \pi)(\psi - \psi_{0,n}) + R_n(\psi, \pi), \end{aligned}$$

where $D_\psi Q_n(\psi_{0,n}, \pi) \in R^{d_\psi}$ is a stochastic generalized first partial-derivative vector and $D_{\psi\psi} Q_n(\psi_{0,n}, \pi) \in R^{d_\psi \times d_\psi}$ is a generalized second partial-derivative matrix that is symmetric and may be stochastic or non-stochastic,

[18] The nth term $a_n(\gamma_n)$ in the sequence of constants $\{a_n(\gamma_n)\}$ actually depends on the entire sequence $\{\gamma_n\}$ because b depends on $\{\gamma_n\}$. For notational simplicity, however, this is not reflected in the notation $a_n(\gamma_n)$.

(ii) the remainder, $R_n(\psi, \pi)$, satisfies

$$\sup_{\psi \in \Psi(\pi): ||\psi - \psi_{0,n}|| \leq \delta_n} \frac{|a_n^2(\gamma_n) R_n(\psi, \pi)|}{(1 + ||a_n(\gamma_n)(\psi - \psi_{0,n})||)^2} = o_{p\pi}(1)$$

for all constants $\delta_n \to 0$, and
(iii) $D_\zeta Q_n(\theta)$ and $D_{\zeta\zeta} Q_n(\theta)$ do not depend on π when $\beta = 0$, where $\theta = (\beta, \zeta, \pi) \in \Theta$, $D_\zeta Q_n(\theta)$ denotes the last d_ζ elements of $D_\psi Q_n(\theta)$, and $D_{\zeta\zeta} Q_n(\theta)$ is the lower $d_\zeta \times d_\zeta$ block of $D_{\psi\psi} Q_n(\theta)$.

Sufficient conditions for Assumption C1 when $Q_n(\theta)$ is a sample average that is smooth in θ are given in Lemma 11.5 in AC1-SM. In this case, $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ are the pointwise partial and second partial derivatives of $Q_n(\theta)$. For the non-smooth sample average case, sufficient conditions are given in Lemma 11.6 in AC1-SM. In this case, $D_\psi Q_n(\theta)$ is a "stochastic derivative" of $Q_n(\theta)$, which typically equals the pointwise derivative for points where the latter exists, and $D_{\psi\psi} Q_n(\theta)$ is the (non-stochastic) second partial derivative of the expected value of $Q_n(\theta)$. For example, this case covers quantile estimators and ML and LS estimators in continuous, but not smooth, threshold autoregressive models, as in Chan and Tsay (1998).

Sufficient conditions for Assumption C1 when $Q_n(\theta)$ is a GMM or MD criterion function, smooth or non-smooth in θ, are given in AC3. In the GMM case, $D_\psi Q_n(\theta)$ is the product of two matrices and a vector: (i) the derivative wrt ψ of the expected value of the moment conditions, (ii) the limit of the GMM weight matrix, and (iii) the sample moment vector. The non-stochastic matrix $D_{\psi\psi} Q_n(\theta)$ is the same as $D_\psi Q_n(\theta)$ except the sample moment vector is replaced by the transpose of the matrix in (i).

If $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ are the pointwise partial and second partial derivatives of $Q_n(\theta)$, then Assumption C1(iii) is implied by Assumption A. When $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ are generalized derivatives, then Assumption C1(iii) is not necessarily implied by Assumption A (because generalized derivatives are not uniquely defined), but in the presence of Assumption A the condition is not restrictive.

Note that Assumption C1 is compatible with semi-parametric estimators.

Example 1 (cont.). The (generalized) first and second derivatives of $Q_n(\theta)$ wrt ψ, which appear in Assumption C1, are the ordinary first and second partial derivatives of the approximation $Q_n^\infty(\theta)$ to $Q_n(\theta)$. Here, $Q_n^\infty(\theta)$ is defined by

$$Q_n^\infty(\theta) = \frac{1}{2}\log\zeta + \frac{1}{2\zeta}n^{-1}\sum_{t=1}^{n}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)^2, \tag{4.7}$$

where the sum over j runs to ∞, rather than to $t-1$. The difference between $Q_n^\infty(\theta)$ and $Q_n(\theta)$ is due to the initial conditions employed: $Q_n^\infty(\theta)$ starts in the infinite past. These differences are shown to be asymptotically negligible using Lemma 11.7 in Appendix A of AC1-SM.

We verify Assumption C1 with

$$\begin{aligned}
D_\psi Q_n(\theta) &= n^{-1}\sum_{t=1}^{n}\rho_{\psi,t}(\theta) = \begin{pmatrix}\rho_{\beta,t}(\theta)\\ \rho_{\zeta,t}(\theta)\end{pmatrix}, \text{ where}\\
\rho_{\beta,t}(\theta) &= -\zeta^{-1}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\sum_{k=0}^{\infty}\pi^k Y_{t-k-1} \text{ and}\\
\rho_{\zeta,t}(\theta) &= -\frac{1}{2}\zeta^{-2}\left(\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)^2 - \zeta\right),
\end{aligned} \tag{4.8}$$

using Lemmas 11.5 and 11.7 in AC1-SM. For brevity, the verification is given in Appendix C of AC1-SM. □

The (generalized) first derivative of $Q_n(\theta)$ wrt ψ is assumed to satisfy:

Assumption C2. (i) $D_\psi Q_n(\theta)$ takes the form

$$D_\psi Q_n(\theta) = n^{-1}\sum_{i=1}^{n} m(W_i, \theta)$$

for some function $m(W_i,\theta) \in R^{d_\psi}$ $\forall\theta \in \Theta_\delta$, for any true parameter $\gamma^* \in \Gamma$.
(ii) $E_{\gamma^*}m(W_i, \psi^*, \pi) = 0$ $\forall\pi \in \Pi$, $\forall i \geq 1$ when the true parameter is γ^* $\forall\gamma^* = (\psi^*, \pi^*, \phi^*) \in \Gamma$ with $\beta^* = 0$.[19]

[19] In some time series examples $D_\psi Q_n(\theta)$ is of the form $n^{-1}\sum_{i=1}^{n} m_i(\theta)$, where $m_i(\theta)$ depends on $\{W_j : \forall 1 \leq j \leq i\}$. Assumption C2 can be relaxed to cover such cases without any changes to the results of the paper. In such cases, Assumption C3 below still can hold provided $\{m_i(\theta) : i \leq n\}$ satisfies a suitable "asymptotic weak dependence" condition, such as near-epoch dependence.

Example 1 (cont.). Assumption C2(i) holds in this example with

$$m(W_i, \theta) = \rho_{\psi,t}(\theta). \tag{4.9}$$

Assumption C2(ii) holds because, for all $\gamma^* \in \Gamma$ with $\beta^* = 0$,

$$\begin{aligned} E_{\gamma^*}\rho_{\beta,t}(\psi^*, \pi) &= -\zeta^{*-1}E_{\gamma^*}\varepsilon_t \sum_{j=0}^{\infty} \pi^j Y_{t-j-1} = 0 \text{ and} \\ E_{\gamma^*}\rho_{\zeta,t}(\psi^*, \pi) &= -(1/2)\zeta^{*-2}(E_{\gamma^*}\varepsilon_t^2 - \zeta^*) = 0 \end{aligned} \tag{4.10}$$

using (4.4) and the definitions of $\rho_{\beta,t}(\theta)$ and $\rho_{\zeta,t}(\theta)$ in (4.8). □

For simplicity, $m(W_i, \theta)$ is abbreviated as $m_i(\theta)$. Define an empirical process $\{G_n(\pi) : \pi \in \Pi\}$ by

$$G_n(\pi) = n^{-1/2} \sum_{i=1}^{n} \left(m_i(\psi_{0,n}, \pi) - E_{\gamma_n} m_i(\psi_{0,n}, \pi) \right). \tag{4.11}$$

The recentered and rescaled (generalized) first derivative of $Q_n(\theta)$ wrt ψ is assumed to satisfy an empirical process CLT:

Assumption C3. Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $G_n(\cdot) \Rightarrow G(\cdot; \gamma_0)$, where $G(\cdot; \gamma_0)$ is a mean zero Gaussian process indexed by $\pi \in \Pi$ with bounded continuous sample paths and some covariance kernel $\Omega(\pi_1, \pi_2; \gamma_0)$ for $\pi_1, \pi_2 \in \Pi$.

Numerous empirical process results in the literature can be used to verify this assumption, including results in Pollard (1984, 1990), Andrews (1994), and van der Vaart and Wellner (1996).

Example 1 (cont.). In this example, the empirical process $\{G_n(\pi) : \pi \in \Pi\}$ in Assumption C3 is defined by

$$\begin{aligned} G_n(\pi) &= n^{-1/2} \sum_{t=1}^{n} \begin{pmatrix} \rho_{\beta,t}(\psi_{0,n}, \pi) \\ \rho_{\zeta,t}(\psi_{0,n}, \pi) \end{pmatrix} - \begin{pmatrix} E_{\gamma_n}\rho_{\beta,t}(\psi_{0,n}, \pi) \\ E_{\gamma_n}\rho_{\zeta,t}(\psi_{0,n}, \pi) \end{pmatrix} \\ &= n^{-1/2} \sum_{t=1}^{n} \begin{pmatrix} -\zeta_n^{-1} Y_t \sum_{k=0}^{\infty} \pi^k Y_{t-k-1} \\ -(1/2)\zeta_n^{-2}(Y_t^2 - \zeta_n) \end{pmatrix} - \begin{pmatrix} -E_{\gamma_n}\zeta_n^{-1} Y_t \sum_{k=0}^{\infty} \pi^k Y_{t-k-1} \\ -E_{\gamma_n}(1/2)\zeta_n^{-2}(Y_t^2 - \zeta_n) \end{pmatrix}. \end{aligned} \tag{4.12}$$

The limit process $\{G(\pi; \gamma_0) : \pi \in \Pi\}$ in Assumption C3 is the mean zero Gaussian process

$$G(\pi;\gamma_0) = \left(\begin{array}{c} \sum_{j=0}^{\infty} \pi^j Z_j \\ (1/2)\zeta_0^{-2}(E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2)^{1/2} Z \end{array} \right), \quad (4.13)$$

where Z, Z_0, $Z_1, ...$ are independent standard normal random variables. The covariance kernel of $G(\pi;\gamma_0)$ is

$$\Omega(\pi_1, \pi_2; \gamma_0) = \left[\begin{array}{cc} (1-\pi_1\pi_2)^{-1} & 0 \\ 0 & (1/4)\zeta_0^{-4} E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2 \end{array} \right]. \quad (4.14)$$

The convergence in Assumption C3 is established using the method in Andrews and Ploberger (1996), see Appendix C of AC1-SM. □

The (generalized) second derivative of $Q_n(\theta)$ wrt ψ is assumed to satisfy:

Assumption C4. (i) Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\sup_{\pi \in \Pi} ||D_{\psi\psi} Q_n(\psi_{0,n}, \pi) - H(\pi;\gamma_0)|| \to_p 0$ for some non-stochastic symmetric $d_\psi \times d_\psi$-matrix-valued function $H(\pi;\gamma_0)$ on $\Pi \times \Gamma$ that is continuous on Π $\forall \gamma_0 \in \Gamma$.
(ii) $\lambda_{\min}(H(\pi;\gamma_0)) > 0$ and $\lambda_{\max}(H(\pi;\gamma_0)) < \infty$ $\forall \pi \in \Pi$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.

Example 1 (cont.). In this example, the quantities $D_{\psi\psi} Q_n(\psi_{0,n}, \pi)$ and $H(\pi;\gamma_0)$ in Assumption C4 are as follows: for $\gamma_0 \in \Gamma$ with $\beta_0 = 0$,

$$\begin{aligned}
& D_{\psi\psi} Q_n(\psi_{0,n}, \pi) = n^{-1} \sum_{t=1}^{n} \rho_{\psi\psi,t}(\psi_{0,n}, \pi) \\
& = n^{-1} \sum_{t=1}^{n} \left[\begin{array}{cc} \rho_{\beta\beta,t}(\psi_{0,n}, \pi) & \rho_{\beta\zeta,t}(\psi_{0,n}, \pi) \\ \rho_{\beta\zeta,t}(\psi_{0,n}, \pi) & \rho_{\zeta\zeta,t}(\psi_{0,n}, \pi) \end{array} \right], \text{ where} \\
& \rho_{\beta\beta,t}(\psi_{0,n}, \pi) = \zeta_n^{-1} \left(\sum_{j=0}^{\infty} \pi^j Y_{t-j-1} \right)^2, \; \rho_{\beta\zeta,t}(\psi_{0,n}, \pi) = \zeta_n^{-2} Y_t \sum_{k=0}^{\infty} \pi^k Y_{t-k-1}, \\
& \rho_{\zeta\zeta,t}(\psi_{0,n}, \pi) = -(1/2)\zeta_n^{-2} + \zeta_n^{-3} Y_t^2, \text{ and} \\
& H(\pi;\gamma_0) = E_{\gamma_0} \rho_{\psi\psi,t}(\psi_0, \pi) = \left[\begin{array}{cc} (1-\pi^2)^{-1} & 0 \\ 0 & (2\zeta_0^2)^{-1} \end{array} \right]. \quad (4.15)
\end{aligned}$$

Assumption C4(i) holds by a uniform LLN, see Appendix C of AC1-SM.

The matrix $H(\pi;\gamma_0)$ satisfies Assumption C4(ii) because $\inf_{\pi \in \Pi}(1-\pi^2)^{-1} = (1 - \max^2\{|\pi_L|, |\pi_U|\})^{-1} > 0$. □

Define the $d_\psi \times d_\beta$-matrix of partial derivatives of the average population moment function wrt the true β value, β^*, to be

$$K_n(\theta;\gamma^*) = n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\beta^{*\prime}}E_{\gamma^*}m(W_i,\theta). \tag{4.16}$$

The domain of the function $K_n(\theta;\gamma^*)$ is $\Theta_\delta \times \Gamma_0$, where $\Gamma_0 = \{\gamma_a = (a\beta,\zeta,\pi,\phi) \in \Gamma : \gamma = (\beta,\zeta,\pi,\phi) \in \Gamma$ with $||\beta|| < \delta$ and $a \in [0,1]\}$ and $\delta > 0$ is as in Assumption B2(iii). The set Γ_0 is not empty by Assumption B2(ii).

Assumption C5. (i) $K_n(\theta;\gamma^*)$ exists $\forall(\theta,\gamma^*) \in \Theta_\delta \times \Gamma_0$, $\forall n \geq 1$.
(ii) For some non-stochastic $d_\psi \times d_\beta$-matrix-valued function $K(\psi_0,\pi;\gamma_0)$, $K_n(\overline{\psi}_n,\pi;\widetilde{\gamma}_n) \to K(\psi_0,\pi;\gamma_0)$ uniformly over $\pi \in \Pi$ for all non-stochastic sequences $\{\overline{\psi}_n\}$ and $\{\widetilde{\gamma}_n\}$ such that $\widetilde{\gamma}_n \in \Gamma$, $\widetilde{\gamma}_n \to \gamma_0 = (0,\zeta_0,\pi_0,\phi_0)$ for some $\gamma_0 \in \Gamma$, $(\overline{\psi}_n,\pi) \in \Theta$, and $\overline{\psi}_n \to \psi_0 = (0,\zeta_0)$.
(iii) $K(\psi_0,\pi;\gamma_0)$ is continuous on Π $\forall\gamma_0 \in \Gamma$ with $\beta_0 = 0$.

Assumption C5 is not restrictive. A set of primitive sufficient conditions for Assumption C5 is given in Appendix A of AC1-SM.

For simplicity, $K(\psi_0,\pi;\gamma_0)$ is abbreviated as $K(\pi;\gamma_0)$. Note that $(\overline{\psi}_n,\widetilde{\gamma}_n)$ in Assumption C5(ii) is in $\Theta_\delta \times \Gamma_0$ for n large.

Assumptions C2, C3, and C5 are used to show that

$$n^{1/2}D_\psi Q_n(\psi_{0,n},\pi) = G_n(\pi) + (K_n(\psi_{0,n},\pi;\gamma_n) + o(1))n^{1/2}\beta_n. \tag{4.17}$$

This leads to the following key result concerning the asymptotic behavior of the normalized (generalized) first derivative $D_\psi Q_n(\psi_{0,n},\pi)$ in the quadratic expansion in Assumption C1:

$$\begin{aligned} & a_n(\gamma_n)D_\psi Q_n(\psi_{0,n},\pi) \\ = \; & [G_n(\pi) + (K_n(\psi_{0,n},\pi;\gamma_n) + o(1))n^{1/2}\beta_n]n^{-1/2}a_n(\gamma_n) \\ \Rightarrow \; & \begin{cases} G(\pi;\gamma_0) + K(\pi;\gamma_0)b & \text{if } n^{1/2}\beta_n \to b \in R^{d_\beta} \\ K(\pi;\gamma_0)\omega_0 & \text{if } ||n^{1/2}\beta_n|| \to \infty \text{ and } \beta_n/||\beta_n|| \to \omega_0, \end{cases} \end{aligned} \tag{4.18}$$

where the convergence results hold under $\{\gamma_n\} \in \Gamma(\gamma_0,0,b)$ with $||b|| < \infty$ and $\{\gamma_n\} \in \Gamma(\gamma_0,\infty,\omega_0)$, respectively, see Lemma 12.1 in Appendix B of AC1-SM. This is the first part of Step 3 in Section 2.

Example 1 (cont.). The matrix $K_n(\theta; \gamma_0)$, which appears in Assumption C5(i), is complicated and, hence, for brevity, is given in (13.34), (13.36), and (13.38) in Appendix C of AC1-SM. Its limit, $K(\pi; \gamma_0)$, which appears in Assumptions C5(ii) and C5(iii) is much simpler and is given by

$$K(\pi; \gamma_0) = \begin{pmatrix} -(1 - \pi_0 \pi)^{-1} \\ 0 \end{pmatrix}. \tag{4.19}$$

See Appendix C of AC1-SM for the verification of Assumption C5. □

Next, we introduce the limits of the concentrated criterion function $Q_n^c(\pi) = Q_n(\widehat{\psi}_n(\pi), \pi)$ (referred to in Step 1 of Section 2 and defined formally in Section 5 below) after suitable normalization. Define a "weighted non-central chi-square" process $\{\xi(\pi; \gamma_0, b) : \pi \in \Pi\}$ and a non-stochastic function $\{\eta(\pi; \gamma_0, \omega_0) : \pi \in \Pi\}$ by

$$\begin{aligned} \xi(\pi; \gamma_0, b) &= -\frac{1}{2} \left(G(\pi; \gamma_0) + K(\pi; \gamma_0)\, b\right)' H^{-1}(\pi; \gamma_0) \left(G(\pi; \gamma_0) + K(\pi; \gamma_0) b\right) \text{ and} \\ \eta(\pi; \gamma_0, \omega_0) &= -\frac{1}{2} \omega_0' K(\pi; \gamma_0)' H^{-1}(\pi; \gamma_0) K(\pi; \gamma_0) \omega_0. \end{aligned} \tag{4.20}$$

The process $\xi(\pi; \gamma_0, b)$ is the limit of $Q_n^c(\pi)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $||b|| < \infty$ and the function $\eta(\pi; \gamma_0, \omega_0)$ is the limit of $Q_n^c(\pi)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $||b|| = \infty$. Note that the components of $\xi(\pi; \gamma_0, b)$ and $\eta(\pi; \gamma_0, \omega_0)$ are from (4.18) and Assumption C4. Under Assumptions C3, C4, and C5(iii), $\{\xi(\pi; \gamma_0, b) : \pi \in \Pi\}$ has bounded continuous sample paths a.s.

Example 1 (cont.). Combining (4.13), (4.15) and (4.19), the stochastic process $\xi(\pi; \gamma_0, b)$ in this example is

$$\begin{aligned} \xi(\pi; \gamma_0, b) = -\frac{1}{2} &\left(G(\pi; \gamma_0) + \begin{pmatrix} -b/(1 - \pi_0 \pi) \\ 0 \end{pmatrix} \right)' \begin{bmatrix} 1 - \pi^2 & 0 \\ 0 & 2\zeta_0^2 \end{bmatrix} \\ &\times \left(G(\pi; \gamma_0) + \begin{pmatrix} -b/(1 - \pi_0 \pi) \\ 0 \end{pmatrix} \right). \end{aligned} \tag{4.21}$$

□

To obtain the asymptotic distribution of $\widehat{\pi}_n$ when $\beta_n = O(n^{-1/2})$ via the continuous mapping theorem, we use the following assumption.

Assumption C6. Each sample path of the stochastic process $\{\xi(\pi; \gamma_0, b) : \pi \in \Pi\}$ in some set $A(\gamma_0, b)$ with $P_{\gamma_0}(A(\gamma_0, b)) = 1$ is minimized over Π at a unique point (which typically depends on the sample path), denoted $\pi^*(\gamma_0, b)$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$, $\forall b$ with $||b|| < \infty$.

In Assumption C6, $\pi^*(\gamma_0, b)$ is random.

We now provide a primitive sufficient condition for Assumption C6 for the case when β is a scalar, i.e., $d_\beta = 1$, which covers many cases of interest. Assumptions C1(iii) and C2 and (4.11) imply that $G(\pi; \gamma_0)$ can be partitioned as $(G_1(\pi)', G_2{}')'$, where $G_1(\pi) \in R^{d_\beta}$, $G_2 \in R^{d_\zeta}$, and G_2 does not depend on π. We partition the covariance kernel $\Omega(\pi_1, \pi_2; \gamma_0)$ in Assumption C3 analogously to $G(\pi; \gamma_0)$ and obtain

$$\Omega(\pi_1, \pi_2; \gamma_0) = \begin{bmatrix} \Omega_{11}(\pi_1, \pi_2; \gamma_0) & \Omega_{12}(\pi_1; \gamma_0) \\ \Omega_{12}(\pi_2; \gamma_0)' & \Omega_{22}(\gamma_0) \end{bmatrix}, \tag{4.22}$$

where $\Omega_{22}(\gamma_0) \in R^{d_\zeta \times d_\zeta}$ does not depend on π. For any $\pi_1, \pi_2 \in \Pi$ and $\pi_1 \neq \pi_2$, $(G_1(\pi_1), G_1(\pi_2), G_2)'$ is normally distributed with mean zero and covariance matrix

$$\Omega_G(\pi_1, \pi_2; \gamma_0) = \begin{bmatrix} \Omega_{11}(\pi_1, \pi_1; \gamma_0) & \Omega_{11}(\pi_1, \pi_2; \gamma_0) & \Omega_{12}(\pi_1; \gamma_0) \\ \Omega_{11}(\pi_2, \pi_1; \gamma_0) & \Omega_{11}(\pi_2, \pi_2; \gamma_0) & \Omega_{12}(\pi_2; \gamma_0) \\ \Omega_{12}(\pi_1; \gamma_0)' & \Omega_{12}(\pi_2; \gamma_0)' & \Omega_{22}(\gamma_0) \end{bmatrix}. \tag{4.23}$$

Typically, the covariance matrix $\Omega_G(\pi_1, \pi_2; \gamma_0)$ takes the form of an outer product, which facilitates the verification of Assumption C6**, as shown in the examples.

Assumption C6**. (i) $d_\beta = 1$ (i.e., β is a scalar).
(ii) $\Omega_G(\pi_1, \pi_2; \gamma_0)$ is positive definite, $\forall \pi_1, \pi_2 \in \Pi$ with $\pi_1 \neq \pi_2$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.

Lemma 4.1. *Assumption* C6** *implies Assumption* C6.

Comment. A slightly more general sufficient condition, Assumption C6*, for Assumption C6 is given in Appendix A of AC1-SM.

Example 1 (cont.). We verify Assumption C6 in this example using Assumption C6** and Lemma 4.1. The covariance kernel $\Omega_G(\pi_1, \pi_2; \gamma_0)$ that appears in Assumption C6** is

$$\Omega_G(\pi_1, \pi_2; \gamma_0) = \begin{bmatrix} (1-\pi_1^2)^{-1} & (1-\pi_1\pi_2)^{-1} & 0 \\ (1-\pi_1\pi_2)^{-1} & (1-\pi_2^2)^{-1} & 0 \\ 0 & 0 & (1/4)\zeta_0^{-4} E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2 \end{bmatrix}. \tag{4.24}$$

It is positive definite because the upper left 2×2 block has determinant equal to zero if and only if $\pi_1 = \pi_2$ by straightforward calculations and $\zeta_0^{-4} E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2 > 0$ by the definitions of Θ^* and Φ^* in (4.3) and (4.4). $\square$

The following assumption is used in the proof of consistency of $\widehat{\pi}_n$ in the "less close, local to $\beta = 0$" case in which $\beta_n \to 0$ and $n^{1/2}||\beta_n|| \to \infty$.

Assumption C7. The non-stochastic function $\eta(\pi; \gamma_0, \omega_0)$ is uniquely minimized over $\pi \in \Pi$ at π_0 $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.

In Assumption C7, the minimizing value π_0 is non-random. Assumption C7 can be verified using the Cauchy-Schwarz inequality or a matrix version of it, see Tripathi (1999), when $K(\pi; \gamma_0)$ and $H(\pi; \gamma_0)$ take proper forms, as in most examples, e.g., see the verification of Assumption C7 for the nonlinear regression example in Appendix E of AC1-SM and the verification of Assumption C7 for GMM estimators in AC3.

Lemma 12.3 in Appendix B of AC1-SM shows that when $\pi = \pi_0$, $K(\pi; \gamma_0) = -H(\pi; \gamma_0) S'_\beta$, where $S_\beta = [I_{d_\beta} : 0] \in R^{d_\beta \times d_\psi}$, whereas this relationship does not hold for $\pi \neq \pi_0$ in general.

Example 1 (cont.). In this example, the function $\eta(\pi; \gamma_0, \omega_0)$ in Assumption C7 is

$$\eta(\pi; \gamma_0, \omega_0) = -\frac{1 - \pi^2}{2(1 - \pi_0 \pi)^2}, \tag{4.25}$$

see Appendix C of AC1-SM. It is uniquely minimized at $\pi = \pi_0$, as required by Assumption C7, because its derivative wrt π is

$$\frac{(\pi - \pi_0)}{(1 - \pi_0 \pi)^3}, \tag{4.26}$$

which is zero for $\pi = \pi_0$, strictly negative for $\pi < \pi_0$, and strictly positive for $\pi > \pi_0$. $\square$

The following technical assumption is used when obtaining a rate of convergence result for $\widehat{\psi}_n$ for sequences $\{\gamma_n\}$ for which $\beta_n \to 0$ and $n^{1/2}||\beta_n|| \to \infty$.

Assumption C8. Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\frac{\partial}{\partial \psi'} E_{\gamma_n} D_\psi Q_n(\psi, \pi_n)|_{\psi = \psi_n} \to H(\pi_0; \gamma_0)$.

By Assumption C4(i), $H(\pi; \gamma_0)$ is the probability limit of $D_{\psi\psi} Q_n(\psi_{0,n}, \pi_n)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$. When $Q_n(\theta)$ is a twice differentiable sample average, $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ are its first and second-order partial derivatives wrt ψ, respectively. One can switch E and ∂ under certain regularity conditions, so that $(\partial/\partial\psi') E_{\gamma_n} D_\psi Q_n(\psi_n, \pi_n)$ is the expectation of $D_{\psi\psi}Q_n(\psi_n, \pi_n)$ in this case. Hence, Assumption C8 can be verified by a uniform LLN and the continuity of $D_{\psi\psi}Q_n(\psi, \pi)$ in ψ. When $Q_n(\theta)$ is non-smooth, one can show that $E_{\gamma_n}D_\psi Q_n(\theta)$ is close to the first-order partial derivative of $Q(\theta; \gamma_0)$ wrt ψ, roughly by switching E_{γ_n} and D_ψ under some regularity conditions, and $D_{\psi\psi}Q_n(\theta)$ is typically taken to be the second-order partial derivative of $Q(\theta; \gamma_0)$ wrt ψ in this case.

Example 1 (cont.). For brevity, the quantity $(\partial/\partial\psi')E_{\gamma_n}D_\psi Q_n(\psi, \pi_n)|_{\psi=\psi_n}$ in Assumption C8 and the verification of Assumption C8 is given in of Appendix C of AC1-SM. □

4.5. Distant from $\beta = 0$ Assumptions

Assumptions D1-D3 below are used to derive asymptotic distributions under sequences of true parameters $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. The "D" denotes that the sequences of true parameters considered are more *distant* from the point of non-identification than are the sequences in the "C" assumptions.

We define a matrix $B(\beta)$ that is used to normalize the (generalized) second-derivative matrix $D^2Q_n(\theta_n)$ of $Q_n(\theta_n)$ (which is introduced in Assumption D1 below) so that it is nonsingular asymptotically, as specified in Assumption D2. Let

$$\begin{aligned} B(\beta) &= \begin{bmatrix} I_{d_\psi} & 0_{d_\psi \times d_\pi} \\ 0_{d_\pi \times d_\psi} & \iota(\beta)I_{d_\pi} \end{bmatrix} \in R^{d_\theta \times d_\theta}, \text{ where} \\ \iota(\beta) &= \begin{cases} \beta & \text{if } \beta \text{ is a scalar} \\ ||\beta|| & \text{if } \beta \text{ is a vector} \end{cases}. \end{aligned} \tag{4.27}$$

We use a different definition of $B(\beta)$ in the scalar and vector β cases because in the scalar case the use of β, rather than $||\beta||$, produces noticeably simpler (but equivalent) formulae, but in the vector case $||\beta||$ is required.

Assumption D1. When the true parameters are $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,
(i) the sample criterion function $Q_n(\theta)$ has a quadratic expansion in θ around θ_n:

$$Q_n(\theta) = Q_n(\theta_n) + DQ_n(\theta_n)'(\theta - \theta_n) + \frac{1}{2}(\theta - \theta_n)D^2Q_n(\theta_n)(\theta - \theta_n) + R_n^*(\theta),$$

where $DQ_n(\theta_n) \in R^{d_\theta}$ is a stochastic generalized first derivative vector and $D^2Q_n(\theta_n) \in R^{d_\theta \times d_\theta}$ is a generalized second derivative matrix that is symmetric and may be stochastic or non-stochastic, and

(ii) the remainder, $R_n^*(\theta)$, satisfies

$$\sup_{\theta\in\Theta_n(\delta_n)} \frac{|nR_n^*(\theta)|}{(1+||n^{1/2}B(\beta_n)(\theta-\theta_n)||)^2} = o_p(1)$$

for all constants $\delta_n \to 0$, where $\Theta_n(\delta_n) = \{\theta \in \Theta : \|\psi - \psi_n\| \leq \delta_n \|\beta_n\|$ and $\|\pi - \pi_n\| \leq \delta_n\}$.

The set $\Theta_n(\delta_n)$ in Assumption D1(ii) is a neighborhood of θ_n whose radius shrinks as the sample size gets larger. In particular, the distance between ψ and ψ_n shrinks faster than $||\beta_n||$ when $\beta_n \to 0$. It is shown below that, under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\widehat{\theta}_n \in \Theta_n(\delta_n)$ with probability that goes to one as $n \to \infty$ for some $\delta_n \to 0$.[20]

The sufficient conditions for Assumption C1 referenced in the previous sub-section also are sufficient for Assumption D1. The quantities $DQ_n(\theta_n)$ and $D^2Q_n(\theta_n)$ take similar forms to $D_\psi Q_n(\psi_{0,n}, \pi)$ and $D_{\psi\psi}Q_n(\psi_{0,n}, \pi)$ (see the discussion following Assumption C1), but involve derivatives wrt θ, not ψ, and hence are not functions of π.

Example 1 (cont.). The matrix $B(\beta)$ for the ARMA example is

$$B(\beta) = \begin{bmatrix} I_2 & 0_2 \\ 0_2' & \beta \end{bmatrix} \in R^{3\times 3}. \tag{4.28}$$

The (generalized) first and second derivatives of $Q_n(\theta)$ wrt θ that appear in Assumption D1 are the ordinary first and second partial derivatives of $Q_n^\infty(\theta)$, defined in (4.7). The first derivatives are

$$\begin{aligned} DQ_n(\theta) &= n^{-1}\sum_{t=1}^{n} \rho_{\theta,t}(\theta) = n^{-1}\sum_{t=1}^{n} (\rho_{\beta,t}(\theta), \rho_{\zeta,t}(\theta), \rho_{\pi,t}(\theta))', \text{ where} \\ \rho_{\pi,t}(\theta) &= -\zeta^{-1}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\beta\sum_{k=0}^{\infty} k\pi^{k-1}Y_{t-k-1}, \end{aligned} \tag{4.29}$$

and $\rho_{\beta,t}(\theta)$ and $\rho_{\zeta,t}(\theta)$ are given in (4.8). For brevity, the second derivatives are given in (13.11)-(13.13) of Appendix C of AC1-SM. Assumption D1 is verified using Lemma 11.5 in AC1-SM, see Appendix C of AC1-SM. □

The next assumption requires good behavior of the (generalized) second derivative of $Q_n(\theta_n)$ after it has been rescaled to eliminate its singularity when β_n converges to

[20] This holds because $\widehat{\theta}_n$ is consistent by Lemma 5.3 below and $\widehat{\psi}_n - \psi_n = o_p(\|\beta_n\|)$ when $\beta_n \to 0$ by Lemma 12.4 in Appendix B of AC1-SM.

zero.

Assumption D2. Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$J_n = B^{-1}(\beta_n) D^2 Q_n(\theta_n) B^{-1}(\beta_n) \to_p J(\gamma_0) \in R^{d_\theta \times d_\theta},$$

where $J(\gamma_0)$ is nonsingular and symmetric.[21]

Example 1 (cont.). Assumption D2 holds in this example with $J(\gamma_0)$ equal to

$$J(\gamma_0) = Diag\left\{ \zeta_0^{-1} E_{\gamma_0} \left(\sum_{j=0}^{\infty} \pi_0^j Y_{t-j-1} \right)^2, \ (2\zeta_0^2)^{-1}, \ \zeta_0^{-1} E_{\gamma_0} \left(\sum_{j=0}^{\infty} j \pi_0^{j-1} Y_{t-j-1} \right)^2 \right\}$$
$$+ \left(\zeta_0^{-1} E_{\gamma_0} \left(\sum_{j=0}^{\infty} \pi_0^j Y_{t-j-1} \right) \sum_{k=0}^{\infty} k \pi_0^{k-1} Y_{t-k-1} \right) \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \qquad (4.30)$$

The proof is given in Appendix C of AC1-SM. □

The following assumption requires the rescaled (generalized) first derivative to satisfy a CLT.

Assumption D3. (i) Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$n^{1/2} B^{-1}(\beta_n) D Q_n(\theta_n) \to_d G^*(\gamma_0) \sim N(0_{d_\theta}, V(\gamma_0)),$$

for some symmetric $d_\theta \times d_\theta$-matrix $V(\gamma_0)$.[22]
(ii) $V(\gamma_0)$ is positive definite $\forall \gamma_0 \in \Gamma$.

Example 1 (cont.). To verify Assumption D3(i) in this example, we have

$$n^{1/2} B^{-1}(\beta_n) D Q_n(\theta_n) = n^{-1/2} \sum_{t=1}^{n} B^{-1}(\beta_n) \rho_{\theta,t}(\theta_n)$$
$$= -n^{-1/2} \sum_{t=1}^{n} \begin{pmatrix} \zeta_n^{-1} \varepsilon_t \sum_{k=0}^{\infty} \pi_n^k Y_{t-k-1} \\ (1/2) \zeta_n^{-2} \left(\varepsilon_t^2 - \zeta_n \right) \\ \zeta_n^{-1} \varepsilon_t \sum_{k=0}^{\infty} k \pi_n^{k-1} Y_{t-k-1} \end{pmatrix} \to_d N(0, V(\gamma_0)), \qquad (4.31)$$

[21] In the vector β case, $J(\gamma_0)$ may depend on ω_0 as well as γ_0.
[22] In the vector β case, $V(\gamma_0)$ may depend on ω_0 as well as γ_0.

where the equalities hold by the definitions in (4.8), (4.28), and (4.29) and the convergence in distribution holds by a triangular array martingale difference CLT.

The matrix $V(\gamma_0)$ equals

$$\begin{aligned} & V(\gamma_0) \\ & = Diag\left\{\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_0^j Y_{t-j-1}\right)^2, \ \frac{E_{\gamma_0}\left(\varepsilon_t^2-\zeta_0\right)^2}{4\zeta_0^4}, \ \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}j\pi_0^{j-1}Y_{t-j-1}\right)^2\right\} \\ & \quad + \left(\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_0^j Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi_0^{k-1}Y_{t-k-1}\right)\times\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{4.32}$$

Note that $J(\gamma_0) = V(\gamma_0)$ if $(2\zeta_0^2)^{-1} = (4\zeta_0^4)^{-1}E_{\gamma_0}\left(\varepsilon_t^2-\zeta_0\right)^2$, which holds when ε_t has a normal distribution.

The verification of the conditions needed for the CLT, the derivation of the form of $V(\gamma_0)$, and the verification of Assumption D3(ii) are given in Appendix C of AC1-SM. □

5. Estimation Results

This section provides the asymptotic results of the paper for the extremum estimator $\widehat{\theta}_n$. Define a concentrated extremum estimator $\widehat{\psi}_n(\pi)$ $(\in \Psi(\pi))$ of ψ for given $\pi \in \Pi$ by

$$Q_n(\widehat{\psi}_n(\pi),\pi) = \inf_{\psi\in\Psi(\pi)} Q_n(\psi,\pi) + o(n^{-1}). \tag{5.1}$$

Let $Q_n^c(\pi)$ denote the concentrated sample criterion function $Q_n(\widehat{\psi}_n(\pi),\pi)$. Define an extremum estimator $\widehat{\pi}_n$ $(\in \Pi)$ by

$$Q_n^c(\widehat{\pi}_n) = \inf_{\pi\in\Pi} Q_n^c(\pi) + o(n^{-1}). \tag{5.2}$$

We assume that the extremum estimator $\widehat{\theta}_n$ in (3.5) can be written as $\widehat{\theta}_n = (\widehat{\psi}_n(\widehat{\pi}_n),\widehat{\pi}_n)$. Note that if (5.1) and (5.2) hold and $\widehat{\theta}_n = (\widehat{\psi}_n(\widehat{\pi}_n),\widehat{\pi}_n)$, then (3.5) automatically holds.

Lemma 5.1. *Suppose Assumptions* A *and* B3 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0)$, *where* $\gamma_0 =$

$(\beta_0, \zeta_0, \pi_0, \phi_0)$,
(a) *when* $\beta_0 = 0$, $\sup_{\pi \in \Pi} ||\widehat{\psi}_n(\pi) - \psi_n|| \to_p 0$ *and* $\widehat{\psi}_n - \psi_n \to_p 0$, *and*
(b) *when* $\beta_0 \neq 0$, $\widehat{\theta}_n - \theta_n \to_p 0$.

For $\gamma_n = (\beta_n, \zeta_n, \pi_n, \phi_n) \in \Gamma$, let $Q_{0,n} = Q_n(\psi_{0,n}, \pi)$, where $\psi_{0,n} = (0, \zeta_n)$ as in Assumption C1. Note that $Q_{0,n}$ does not depend on π by Assumption A.

Lemma 5.2. *Suppose Assumptions* A, B1-B3, *and* C1-C5 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$,
(a) *when* $||b|| < \infty$, $n(Q_n^c(\cdot) - Q_{0,n}) \Rightarrow \xi(\cdot; \gamma_0, b)$, *and*
(b) *when* $||b|| = \infty$ *and* $\beta_n/||\beta_n|| \to \omega_0$ *for some* $\omega_0 \in R^{d_\beta}$ *with* $||\omega_0|| = 1$,
$||\beta_n||^{-2}(Q_n^c(\pi) - Q_{0,n}) \to_p \eta(\pi; \gamma_0, \omega_0)$ *uniformly over* $\pi \in \Pi$.

Define the Gaussian process $\{\tau(\pi; \gamma_0, b) : \pi \in \Pi\}$ by

$$\tau(\pi; \gamma_0, b) = -H^{-1}(\pi; \gamma_0)(G(\pi; \gamma_0) + K(\pi; \gamma_0)b) - (b, 0_{d_\zeta}), \tag{5.3}$$

where $(b, 0_{d_\zeta}) \in R^{d_\psi}$. Note that, by (4.20) and (5.3), $\xi(\pi; \gamma_0, b) = -(1/2)(\tau(\pi; \gamma_0, b) + (b, 0_{d_\zeta}))' H(\pi; \gamma_0)(\tau(\pi; \gamma_0, b) + (b, 0_{d_\zeta}))$. Let

$$\pi^*(\gamma_0, b) = \arg\min_{\pi \in \Pi} \xi(\pi; \gamma_0, b). \tag{5.4}$$

Theorem 5.1. *Suppose Assumptions* A, B1-B3, *and* C1-C6 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $||b|| < \infty$,
(a) $\begin{pmatrix} n^{1/2}(\widehat{\psi}_n - \psi_n) \\ \widehat{\pi}_n \end{pmatrix} \to_d \begin{pmatrix} \tau(\pi^*(\gamma_0, b); \gamma_0, b) \\ \pi^*(\gamma_0, b) \end{pmatrix}$ *and*
(b) $n\left(Q_n(\widehat{\theta}_n) - Q_{0,n}\right) \to_d \inf_{\pi \in \Pi} \xi(\pi; \gamma_0, b)$.

Comments. 1. Define the Gaussian process $\{\tau_\beta(\pi; \gamma_0, b) : \pi \in \Pi\}$ by

$$\tau_\beta(\pi; \gamma_0, b) = S_\beta \tau(\pi; \gamma_0, b) + b, \tag{5.5}$$

where $S_\beta = [I_{d_\beta} : 0_{d_\beta \times d_\zeta}]$ is the $d_\beta \times d_\psi$ selector matrix that selects β out of ψ. The asymptotic distribution of $n^{1/2}\widehat{\beta}_n$ (without centering at β_n) under $\Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$ is given by $\tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)$. This quantity appears in the asymptotic distributions of t statistics below.

2. Assumption C6 is not needed for Theorem 5.1(b).

Lemma 5.3. *Suppose Assumptions* A, B1-B3, C1-C5, *and* C7 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,
(a) $\widehat{\pi}_n - \pi_n \to_p 0$ *and* (b) $\widehat{\psi}_n - \psi_n \to_p 0$.

Theorem 5.2. *Suppose Assumptions* A, B1-B3, C1-C5, C7, C8, *and* D1-D3 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,
(a) $n^{1/2}B(\beta_n)(\widehat{\theta}_n - \theta_n) \to_d -J^{-1}(\gamma_0)G^*(\gamma_0) \sim N(0_{d_\theta}, J^{-1}(\gamma_0)V(\gamma_0)J^{-1}(\gamma_0))$ *and*
(b) $n(Q_n(\widehat{\theta}_n) - Q_n(\theta_n)) \to_d -\frac{1}{2}G^*(\gamma_0)'J^{-1}(\gamma_0)G^*(\gamma_0)$.

Example 1 (cont.). In this example, the components of the stochastic processes $\xi(\pi;\gamma_0,b)$ and $\tau(\pi;\gamma_0,b)$, that appear in the asymptotic results in this section, are

$$\begin{aligned} H(\pi;\gamma_0) &= Diag((1-\pi^2)^{-1}, (2\zeta_0^2)^{-1}), \\ K(\pi;\gamma_0) &= (-(1-\pi_0\pi)^{-1}, 0)', \\ \Omega(\pi_1,\pi_2;\gamma_0) &= Diag((1-\pi_1\pi_2)^{-1}, (1/4)\zeta_0^{-4}E_{\gamma_0}\left(\varepsilon_t^2 - \zeta_0\right)^2), \end{aligned} \tag{5.6}$$

and $G(\pi;\gamma_0)$ is a mean zero Gaussian process with covariance kernel $\Omega(\pi_1,\pi_2;\gamma_0)$. In addition, $J(\gamma_0)$ and $V(\gamma_0)$ are defined in (4.30) and (4.32), respectively. In addition,

$$\tau_\beta(\pi;\gamma_0,b) = -(1-\pi^2)\left(\sum_{j=0}^{\infty}\pi^j Z_j - (1-\pi_0\pi)^{-1}b\right). \ \square \tag{5.7}$$

6. t Confidence Intervals and Tests

In this section, we consider a confidence interval (CI) for a real-valued function $r(\theta)$ of θ by inverting a t test of the hypotheses $H_0 : r(\theta) = v$ for $v \in r(\Theta)$. We also consider t tests of H_0. We determine the asymptotic size of standard t CI's. We introduce robust t CI's whose asymptotic size is guaranteed to equal their nominal size. For brevity, results for Wald CS's and tests for vector-valued functions $r(\theta)$ are given in AC3.

6.1. t Statistics

The t statistic is defined as follows. Let

$$\Sigma(\gamma_0) = J^{-1}(\gamma_0)V(\gamma_0)J^{-1}(\gamma_0) \text{ and } \widehat{\Sigma}_n = \widehat{J}_n^{-1}\widehat{V}_n\widehat{J}_n^{-1}, \tag{6.1}$$

where $\widehat{J}_n$ and $\widehat{V}_n$ are estimators of $J(\gamma_0)$ and $V(\gamma_0)$ that do not depend on the nuisance parameter ϕ.

The t statistic takes the form

$$T_n(v) = \frac{n^{1/2}(r(\widehat{\theta}_n) - v)}{(r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)')^{1/2}}, \tag{6.2}$$

where $r_\theta(\theta) = (\partial/\partial\theta')r(\theta) = [r_\psi(\theta) : r_\pi(\theta)] \in R^{d_r \times d_\theta}$, $r_\psi(\theta) = (\partial/\partial\psi')r(\theta) \in R^{d_r \times d_\psi}$, and $r_\pi(\theta) = (\partial/\partial\pi')r(\theta) \in R^{d_r \times d_\pi}$.

Although this definition of the t statistic involves $B^{-1}(\widehat{\beta}_n)$, it is the same as the standard definition used in practice. By Theorem 5.2(a), when $\beta_0 \neq 0$, $B^{-1}(\beta_0)\Sigma(\gamma_0)B^{-1}(\beta_0)$ is the asymptotic covariance matrix of $\widehat{\theta}_n$. In the t statistic, the asymptotic covariance is replaced by the estimator $B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)$. The same form of the t statistic is used under all sequences of true parameters $\gamma_n \in \Gamma(\gamma_0)$.

In the results below, we consider the behavior of the t statistic when the null hypothesis holds. Thus, under a sequence $\{\gamma_n\}$, we consider the sequence of null hypotheses $H_0 : r(\theta) = v_n$, where v_n equals $r(\theta_n)$ and $\gamma_n = (\theta_n, \phi_n)$. We employ the following notational simplification:

$$T_n = T_n(v_n), \text{ where } v_n = r(\theta_n). \tag{6.3}$$

6.2. Function of Interest

Let d_r denote the dimension of $r(\theta)$. Here, $d_r = 1$. (In AC3, $d_r > 1$ is considered.)

The function of interest, $r(\theta)$, satisfies the following assumption.

Assumption R1. (i) $r(\theta)$ is continuously differentiable on Θ.
(ii) $r_\theta(\theta)$ is full row rank d_r $\forall \theta \in \Theta$.
(iii) $rank(r_\pi(\theta)) = d_\pi^*$ for some constant $d_\pi^* \leq \min(d_r, d_\pi)$ $\forall \theta \in \Theta_\delta = \{\theta \in \Theta : ||\beta|| < \delta\}$ for some $\delta > 0$.

A sufficient condition for Assumption R1 is: $r(\theta) = R_1'\theta$, where $R_1 \in R^{d_\theta}$ and $R_1 \neq 0$.

6.3. Variance Matrix Estimators

The estimators of the components of the asymptotic variance matrix are assumed to satisfy the following assumptions. Two forms are given for Assumption V1 that follows.

The first applies when β is a scalar and the second applies when β is a vector. The reason for the difference is that the normalizing matrix $B(\beta)$ is different in these two cases.

When β is a scalar, let $J(\theta;\gamma_0)$ and $V(\theta;\gamma_0)$ for $\theta \in \Theta$ be some non-stochastic $d_\theta \times d_\theta$ matrix-valued functions such that $J(\theta_0;\gamma_0) = J(\gamma_0)$ and $V(\theta_0;\gamma_0) = V(\gamma_0)$, where $J(\gamma_0)$ and $V(\gamma_0)$ are as in Assumptions D2 and D3. Let

$$\Sigma(\theta;\gamma_0) = J^{-1}(\theta;\gamma_0)V(\theta;\gamma_0)J^{-1}(\theta;\gamma_0) \text{ and } \Sigma(\pi;\gamma_0) = \Sigma(\psi_0,\pi;\gamma_0). \quad (6.4)$$

Let $\Sigma_{\beta\beta}(\pi;\gamma_0)$ denote the upper left (1,1) element of $\Sigma(\pi;\gamma_0)$.

Assumption V1 below applies when β is a scalar.

Assumption V1 (scalar $\boldsymbol{\beta}$). (i) $\widehat{J}_n = \widehat{J}_n(\widehat{\theta}_n)$ and $\widehat{V}_n = \widehat{V}_n(\widehat{\theta}_n)$ for some (stochastic) $d_\theta \times d_\theta$ matrix-valued functions $\widehat{J}_n(\theta)$ and $\widehat{V}_n(\theta)$ on Θ that satisfy $\sup_{\theta\in\Theta} ||\widehat{J}_n(\theta) - J(\theta;\gamma_0)|| \to_p 0$ and $\sup_{\theta\in\Theta} ||\widehat{V}_n(\theta) - V(\theta;\gamma_0)|| \to_p 0$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$.
(ii) $J(\theta;\gamma_0)$ and $V(\theta;\gamma_0)$ are continuous in θ on Θ $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.
(iii) $\lambda_{\min}(\Sigma(\pi;\gamma_0)) > 0$ and $\lambda_{\max}(\Sigma(\pi;\gamma_0)) < \infty$ $\forall \pi \in \Pi$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.

When β is a vector, i.e., $d_\beta > 1$, we reparameterize β as $(||\beta||, \omega)$, where $\omega = \beta/||\beta||$ if $\beta \neq 0$ and by definition $\omega = 1_{d_\beta}/||1_{d_\beta}||$ with $1_{d_\beta} = (1, ..., 1) \in R^{d_\beta}$ if $\beta = 0$. Correspondingly, θ is reparameterized as $\theta^+ = (||\beta||, \omega, \zeta, \pi)$. Let $\Theta^+ = \{\theta^+ : \theta^+ = (||\beta||, \beta/||\beta||, \zeta, \pi), \theta \in \Theta\}$. Let $\widehat{\theta}_n^+$ and θ_0^+ be the counterparts of $\widehat{\theta}_n$ and θ_0 after reparametrization.

When β is a vector, let $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ denote some non-stochastic $d_\theta \times d_\theta$ matrix-valued functions such that $J(\theta_0^+;\gamma_0) = J(\gamma_0)$ and $V(\theta_0^+;\gamma_0) = V(\gamma_0)$. Let

$$\begin{aligned}\Sigma(\theta^+;\gamma_0) &= J^{-1}(\theta^+;\gamma_0)V(\theta^+;\gamma_0)J^{-1}(\theta^+;\gamma_0) \text{ and}\\ \Sigma(\pi,\omega;\gamma_0) &= \Sigma(||\beta_0||,\omega,\zeta_0,\pi;\gamma_0). \quad (6.5)\end{aligned}$$

Let $\Sigma_{\beta\beta}(\pi,\omega;\gamma_0)$ denote the upper left $d_\beta \times d_\beta$ sub-matrix of $\Sigma(\pi,\omega;\gamma_0)$.

Assumption V1 below applies when β is a vector.

Assumption V1 (vector $\boldsymbol{\beta}$). (i) $\widehat{J}_n = \widehat{J}_n(\widehat{\theta}_n^+)$ and $\widehat{V}_n = \widehat{V}_n(\widehat{\theta}_n^+)$ for some (stochastic) $d_\theta \times d_\theta$ matrix-valued functions $\widehat{J}_n(\theta^+)$ and $\widehat{V}_n(\theta^+)$ on Θ^+ that satisfy $\sup_{\theta^+\in\Theta^+} ||\widehat{J}_n(\theta^+) - J(\theta^+;\gamma_0)|| \to_p 0$ and $\sup_{\theta^+\in\Theta^+} ||\widehat{V}_n(\theta^+) - V(\theta^+;\gamma_0)|| \to_p 0$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with

$||b|| < \infty$.[23]
(ii) $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ are continuous in θ^+ on Θ^+ $\forall\gamma_0 \in \Gamma$ with $\beta_0 = 0$.
(iii) $\lambda_{\min}(\Sigma(\pi,\omega;\gamma_0)) > 0$ and $\lambda_{\max}(\Sigma(\pi,\omega;\gamma_0)) < \infty$ $\forall\pi \in \Pi$, $\forall\omega \in R^{d_\beta}$ with $||\omega|| = 1$, $\forall\gamma_0 \in \Gamma$ with $\beta_0 = 0$.
(iv) $P(\tau_\beta(\pi^*(\gamma_0,b);\gamma_0,b) = 0) = 0$ $\forall\gamma_0 \in \Gamma$ with $\beta_0 = 0$ and $\forall b$ with $||b|| < \infty$.[24]

The following assumption applies with both scalar and vector β.

Assumption V2. Under $\Gamma(0,\infty,\omega_0)$, $\widehat{J}_n \to_p J(\gamma_0)$ and $\widehat{V}_n \to_p V(\gamma_0)$.

Example 1 (cont.). In this example, we estimate $J(\gamma_0)$ and $V(\gamma_0)$ by $\widehat{J}_n = \widehat{J}_n(\widehat{\theta}_n)$ and $\widehat{V}_n = \widehat{V}_n(\widehat{\theta}_n)$, respectively, where

$$\widehat{J}_n(\theta) = Diag\left\{\zeta^{-1}n^{-1}\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1}\pi^j Y_{t-j-1}\right)^2,\ (2\zeta^2)^{-1},\ \zeta^{-1}n^{-1}\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1} j\pi^{j-1}Y_{t-j-1}\right)^2\right\}$$
$$+\left(\zeta^{-1}n^{-1}\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1}\pi^j Y_{t-j-1}\right)\sum_{k=0}^{t-1}k\pi^{k-1}Y_{t-k-1}\right) \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (6.6)$$

and $\widehat{V}_n(\theta)$ equals $\widehat{J}_n(\theta)$ but with its $(2,2)$ element, $(2\zeta^2)^{-1}$, replaced by

$$(4\zeta^2)^{-1}n^{-1}\sum_{t=1}^{n}\left(\left(Y_t - \beta\sum_{j=0}^{t-1}\pi^j Y_{t-j-1}\right)^2 - \zeta\right)^2. \quad (6.7)$$

For hypotheses and CI's that involve only β and/or π, the $(2,2)$ elements of $\widehat{J}_n$ and $\widehat{V}_n$ are not needed. In such cases, the matrices $\widehat{J}_n$ and $\widehat{V}_n$ with their second rows and columns deleted are the same. For Assumptions V1 and V2 to hold for the quantity in (6.7) more moments need to be assumed on ε_t. Specifically, in Φ (defined in (4.4)), the condition $E_\phi|\xi_t|^{4+\delta_2} \leq K$ needs to be replaced by $E_\phi|\xi_t|^{8+\delta_2} \leq K$ for the proof to go through. This condition is only needed for hypotheses and CI's that involve the innovation variance ζ.

For brevity, the quantities $J(\theta;\gamma_0)$ and $V(\theta;\gamma_0)$ in Assumption V1 (scalar β) are given in (13.57) and (13.58) of AC1-SM and Assumptions V1 (scalar β) and V2 are

[23] The functions $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ do not depend on ω_0, only γ_0.
[24] Assumption V1 (vector β) differs from Assumption V1 (scalar β) because in the vector β case Assumption V1(ii) (scalar β) (i.e., continuity in θ) often fails, but Assumption V1(ii) (vector β) (i.e., continuity in θ^+) holds.

verified in Appendix C of AC1-SM. □

6.4. Asymptotic Distribution of the t Statistic

Next, we provide the asymptotic distribution of the t statistic under H_0. Define

$$T_\psi(\pi;\gamma_0,b)=\frac{r_\psi(\pi)\tau(\pi;\gamma_0,b)}{(r_\psi(\pi)\overline{\Sigma}_{\psi\psi}(\pi;\gamma_0,b)r_\psi(\pi)')^{1/2}}, \tag{6.8}$$

where $r_\psi(\pi)=r_\psi(\psi_0,\pi)\in R^{1\times d_\psi}$, $\tau(\pi;\gamma_0,b)\in R^{d_\psi}$, $\overline{\Sigma}_{\psi\psi}(\pi;\gamma_0,b)$ is the upper left $d_\psi\times d_\psi$ block of $\overline{\Sigma}(\pi;\gamma_0,b)$,

$$\overline{\Sigma}(\pi;\gamma_0,b) = \begin{cases}\Sigma(\pi;\gamma_0) & \text{if } \beta \text{ is a scalar}\\ \Sigma(\pi,\omega^*(\pi;\gamma_0,b);\gamma_0) & \text{if } \beta \text{ is a vector,}\end{cases}$$
$$\omega^*(\pi;\gamma_0,b) = \tau_\beta(\pi;\gamma_0,b)/||\tau_\beta(\pi;\gamma_0,b)||, \tag{6.9}$$

$\Sigma(\pi;\gamma_0)$, $\Sigma(\pi,\omega;\gamma_0)$, and $\tau_\beta(\pi;\gamma_0,b)$ are defined in (6.4), (6.5), and (5.5), respectively.

Define

$$T_\pi(\pi;\gamma_0,b)=\frac{||\tau_\beta(\pi;\gamma_0,b)||(r(\psi_0,\pi)-r(\psi_0,\pi_0))}{(r_\pi(\pi)\overline{\Sigma}_{\pi\pi}(\pi;\gamma_0,b)r_\pi(\pi)')^{1/2}}, \tag{6.10}$$

where $\overline{\Sigma}_{\pi\pi}(\pi;\gamma_0,b)$ is the lower right $d_\pi\times d_\pi$ block of $\overline{\Sigma}(\pi;\gamma_0,b)$, and $r_\pi(\pi)=r_\pi(\psi_0,\pi)$.

The following theorem provides the asymptotic null distribution of the t statistic for a scalar restriction. (The null holds by the definition $T_n=T_n(v_n)$ in (6.3).)

Theorem 6.1. *Suppose Assumptions* A, B1-B3, C1-C8, D1-D3, R1, *and* V1-V2 *hold and* $d_r=1$.
(a) *Under* $\{\gamma_n\}\in\Gamma(\gamma_0,0,b)$ *with* $||b||<\infty$ *and* $d_\pi^*=0$, $T_n\to_d T_\psi(\pi^*(\gamma_0,b);\gamma_0,b)$.
(b) *Under* $\{\gamma_n\}\in\Gamma(\gamma_0,0,b)$ *with* $||b||<\infty$ *and* $d_\pi^*=1$, $T_n\to_d T_\pi(\pi^*(\gamma_0,b);\gamma_0,b)$.
(c) *Under* $\{\gamma_n\}\in\Gamma(\gamma_0,\infty,\omega_0)$, $T_n\to_d N(0,1)$.

Comments. 1. When $d_\pi^*=0$, the scalar restriction only involves ψ by Assumption R1(iii). When $d_\pi^*=1$, the restriction involves π and possibly ψ. However, the randomness in $\widehat{\psi}_n$ is dominated by that in $\widehat{\pi}_n$ under the conditions of Theorem 6.1(b) because $\widehat{\psi}_n$ is consistent but $\widehat{\pi}_n$ is not. In consequence, the asymptotic distribution in Theorem 6.1(b) is as if the restriction is only on π.

2. To establish the asymptotic distribution of the t statistic we consider a rotation of $r(\widehat{\theta}_n)$ and $r_\theta(\widehat{\theta}_n)$ by a matrix $A(\widehat{\theta}_n)$. The rotation is designed to separate the effects of the randomness in $\widehat{\psi}_n$ and $\widehat{\pi}_n$, which have different rates of convergence for some sequences $\{\gamma_n\}$. Similar rotations are carried out in the analysis of partially-identified models in Sargan (1983) and Phillips (1989), in the nonstationary time series literature, e.g., see Park and Phillips (1988), and in the GMM analysis in Antoine and Renault (2009, 2010).

Example 1 (cont.). In this example, the asymptotic null distribution of the t statistic for tests concerning the MA parameter π is determined by Theorem 6.1(b) to be as follows. Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $|b| < \infty$, it is the distribution of

$$\begin{aligned} T_\pi(\pi^*; \gamma_0, b) &= \frac{|\tau_\beta(\pi^*; \gamma_0, b)|(\pi^* - \pi_0)}{\left(\overline{\Sigma}_{\pi\pi}(\pi^*; \gamma_0, b)\right)^{1/2}} \\ &= \frac{\left|\sum_{j=0}^{\infty} \pi^{*j} Z_j - (1 - \pi_0 \pi^*)^{-1} b\right| (1 - \pi^{*2})(\pi^* - \pi_0)}{(\Sigma_{\pi\pi}(\pi^*)_{22})^{1/2}}, \end{aligned} \tag{6.11}$$

where π^* abbreviates $\pi^*(\gamma_0, b)$, $\{Z_j : j \geq 0\}$ are i.i.d. $N(0,1)$ random variables,

$$\begin{aligned} \pi^*(\gamma_0, b) &= \underset{\pi \in \Pi}{\arg\min}\ \xi(\pi; \gamma_0, b) \\ &= \underset{\pi \in \Pi}{\arg\min}\ -\frac{1}{2}\left(\sum_{j=0}^{\infty} \pi^j Z_j - (1 - \pi_0 \pi)^{-1} b\right)^2 (1 - \pi^2), \\ \Sigma_{\pi\pi}(\pi) &= \begin{bmatrix} \sum_{j=0}^{\infty} \pi^{2j} & \sum_{j=0}^{\infty} j\pi^{2j-1} \\ \sum_{j=0}^{\infty} j\pi^{2j-1} & \sum_{j=0}^{\infty} j^2 \pi^{2j-2} \end{bmatrix}^{-1}, \end{aligned} \tag{6.12}$$

and $\Sigma_{\pi\pi}(\pi)_{22}$ denotes the $(2,2)$ element of $\Sigma_{\pi\pi}(\pi)$. The second equality in (6.12) holds using the expression for $\xi(\pi; \gamma_0, b)$ in this example given in (4.21) plus simplifications based on (4.13), (4.15), and (4.19).[25] The third equality in (6.11) uses (5.7) and the equality $\overline{\Sigma}_{\pi\pi}(\pi; \gamma_0, b) = \Sigma_{\pi\pi}(\pi)_{22}$, which holds using the expressions for $J(\theta; \gamma_0)$ and $V(\theta; \gamma_0)$ in (13.57) and (13.58) of AC1-SM and some calculations. The limit distribution in (6.11) only depends on b and π_0.

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, the t statistic for the MA parameter π has a $N(0,1)$ asymptotic null distribution by Theorem 6.1(c).

The asymptotic null distribution of the t statistic for tests concerning the AR parameter $\rho = \pi + \beta$ is the same as in (6.11) except that the denominator $(\Sigma_{\pi\pi}(\pi)_{22})^{1/2}$ is

[25] The equality in (6.12) uses the block diagonality of $H(\pi; \gamma_0)$ in (4.15) and the fact that the second element of $G(\pi; \gamma_0)$ in (4.13) does not depend on π.

replaced by $(1_2' \Sigma_{\pi\pi}(\pi) 1_2)^{1/2}$, where $1_2 = (1,1)'$.

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, the t statistic for the AR parameter $\rho = \pi + \beta$ has a $N(0,1)$ asymptotic null distribution by Theorem 6.1(c). $\square$

6.5. Asymptotic Size of Standard t Confidence Intervals

Now, we establish the asymptotic size of a standard confidence interval (CI) obtained by inverting a t statistic. The usual symmetric two-sided t CI takes the form

$$CI_{t,n} = \{v : |T_n(v)| \leq z_{1-\alpha/2}\}, \tag{6.13}$$

where the t statistic $T_n(v)$ is as in (6.2), $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution, and $1-\alpha$ is the nominal size of the CI. Standard upper and lower one-sided t CI's are obtained by replacing $|T_n(v)|$ with $T_n(v)$ and $-T_n(v)$, respectively, and using $z_{1-\alpha}$ as the critical value.

The asymptotic size of the CI above is established by verifying the high-level conditions in Andrews, Cheng, and Guggenberger (2009), hereafter ACG. In particular, assumptions in ACG require the asymptotic distribution of T_n, which abbreviates $T_n(r(\theta_n))$, under drifting sequences of true parameters. Such asymptotic distributions are given in Theorem 6.1.

Define

$$\begin{aligned} h &= (b, \gamma_0), \\ H &= \{h = (b, \gamma_0) : ||b|| < \infty, \gamma_0 \in \Gamma \text{ with } \beta_0 = 0\}, \text{ and} \\ T(h) &= \begin{cases} T_\psi(\pi^*(\gamma_0, b); \gamma_0, b) & \text{if } d_\pi^* = 0 \\ T_\pi(\pi^*(\gamma_0, b); \gamma_0, b) & \text{if } d_\pi^* = 1 \end{cases} \end{aligned} \tag{6.14}$$

for $||b|| < \infty$. As defined, $T(h)$ is the asymptotic distribution of T_n under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $||b|| < \infty$ given in Theorem 6.1(a) or 6.1(b) depending on the rank of $r_\pi(\theta)$, which is denoted by d_π^*. Only one of the cases applies for any particular parameter of interest $r(\theta)$ and it is known which applies.

Let $c_{|t|,1-\alpha}(h)$, $c_{t,1-\alpha}(h)$, and $c_{-t,1-\alpha}(h)$ denote the $1-\alpha$ quantile of $|T(h)|$, $T(h)$, and $-T(h)$ for $h \in H$.

As in (3.8), $AsySz$ denotes the asymptotic size of a CI of nominal level $1-\alpha$. The asymptotic size results use the following distribution function (df) continuity assumption, which typically is not restrictive.

Assumption V3. The df of $T(h)$ is continuous at $z_{\alpha/2}$, z_α, $z_{1-\alpha}$, and $z_{1-\alpha/2}$ $\forall h \in H$.

Theorem 6.2. *Suppose Assumptions* A, B1-B3, C1-C8, D1-D3, R1, *and* V1-V3 *hold and* $d_r = 1$. *The standard nominal* $1-\alpha$ *symmetric two-sided, upper one-sided, and lower one-sided* t *CI's have* $AsySz = \min\{\inf_{h\in H} P(|T(h)| \leq z_{1-\alpha/2}), 1-\alpha\}$, $\min\{\inf_{h\in H} P(T(h) \leq z_{1-\alpha}), 1-\alpha\}$, *and* $\min\{\inf_{h\in H} P(-T(h) \leq z_{1-\alpha}), 1-\alpha\}$, *respectively.*

7. Robust Confidence Intervals

In this section, we construct robust CI's for $r(\theta)$ that have correct asymptotic size. A robust CI is obtained by inverting a test statistic, denoted here generically by $\mathcal{T}_n$, using a robust critical value that differs from a standard strong-identification critical value (such as a normal or $\chi^2_{d_r}$ quantile). The robust critical value can be data dependent. The test statistic $\mathcal{T}_n$ can be the t statistic defined in (6.13), the QLR statistic analyzed in AC2, the Wald statistic analyzed in AC3, or some other statistic.

A robust critical value takes into account the fact that the test statistic, $\mathcal{T}_n$, has a non-standard asymptotic distribution under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$. As a result, a larger critical value often is required under weak identification, i.e., $||b|| < \infty$, than under semi-strong or strong identification, i.e., $||b|| = \infty$.

A simple robust critical value is the "least-favorable" (LF) critical value that is large enough for all identification categories. This yields a CI with correct asymptotic size, but one that typically is overly long and is not as informative as desirable when the model is strongly identified.

In consequence, we introduce data-dependent critical values that improve upon the LF critical value by using an identification-category-selection (ICS) procedure in the construction of the critical value. Two methods are considered: type 1 and type 2. The first is relatively simple. The second has preferable statistical properties, but is more intensive computationally.

We also introduce versions of these robust critical values that (i) impose the known null hypothesis value and (ii) plug-in consistent estimators of consistently estimable nuisance parameters in the formulae for the robust critical values. We recommend employing combined null-imposed/plug-in versions of the robust critical values whenever possible because they yield the smallest critical values and still deliver asymptotically correct size. However, they may be more burdensome computationally than other versions of the robust critical values.

7.1. Least Favorable Critical Values

Let $\mathcal{T}(h)$ denote the asymptotic distribution of $\mathcal{T}_n$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, where $h = (b, \gamma_0) \in H$ and h and H are defined in (6.14). Let $c_{\mathcal{T},1-\alpha}(h)$ denote the $1-\alpha$ quantile of $\mathcal{T}(h)$ for $h \in H$. For example, when $\mathcal{T}_n$ is the two-sided t statistic $|T_n|$ of Section 6, then $\mathcal{T}(h)$ and $c_{\mathcal{T},1-\alpha}(h)$ equal $|T(h)|$ and $c_{|t|,1-\alpha}(h)$, respectively.

Under semi-strong and strong identification, i.e., $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\mathcal{T}_n$ is assumed to have a standard asymptotic distribution, such as the standard normal or chi-squared distribution. Let $c_{\mathcal{T},1-\alpha}(\infty)$ denote the $1-\alpha$ quantile of this distribution.

The LF critical value is

$$c^{LF}_{\mathcal{T},1-\alpha} = \max\{\sup_{h \in H} c_{\mathcal{T},1-\alpha}(h), c_{\mathcal{T},1-\alpha}(\infty)\}. \tag{7.1}$$

The LF critical value can be improved (i.e., made smaller) by exploiting the knowledge of the null hypothesis value of $r(\theta)$. For example, if the null hypothesis specifies the value of π to be 3, then the supremum in (7.1) does not need to be taken over all $h \in H$, only over the h values for which $\pi = 3$. We call such a critical value a null-imposed (NI) LF critical value. Using a NI-LF critical value increases the computational burden because a different critical value is employed for each null hypothesis value.

To be precise, let

$$H(v) = \{h = (b, \gamma_0) \in H : ||b|| < \infty, r(\theta_0) = v\}, \tag{7.2}$$

where $\gamma_0 = (\theta_0, \phi_0)$. By definition, $H(v)$ is the subset is H that is consistent with the null hypothesis $H_0 : r(\theta_0) = v$, where θ_0 denotes the true value. The NI-LF critical value, denoted $c^{LF}_{\mathcal{T},1-\alpha}(v)$, is defined by replacing H by $H(v)$ in (7.1) when the null hypothesis value is $r(\theta_0) = v$. Note that v takes values in the set $V_r = \{v_0 : r(\theta_0) = v_0$ for some $h = (b, \gamma_0) \in H\}$.[26]

When part of γ is unknown under H_0 but can be consistently estimated, then a *plug-*

[26] When $r(\theta) = \beta$ and the null hypothesis imposes that $\beta = v$, the parameter b can be imposed to equal $n^{1/2}v$. In this case, $H(v) = H_n(v) = \{h = (b, \gamma_0) \in H : b = n^{1/2}v\}$. The asymptotic size results given below for NI LF CI's and NI robust CI's hold in this case.

in LF (or plug-in NI-LF) critical value can be used that has correct size asymptotically and is smaller than the LF (or NI-LF) critical value. The plug-in critical value replaces elements of γ with consistent estimators in the formulae in (7.1) and the supremum over H (or $H(v)$) is reduced to a supremum over the resulting subset of H, denoted $\widehat{H}_n$, for which the consistent estimators appear in each vector γ. For example, if ζ is consistently estimated by $\widehat{\zeta}_n$, then H is replaced by

$$\widehat{H}_n = \{h = (b, \gamma) \in H : \gamma = (\beta, \widehat{\zeta}_n, \pi, \phi)\} \tag{7.3}$$

or $H(v)$ is replaced by $H(v) \cap \widehat{H}_n$. Note that the parameter b is not consistently estimable, so it cannot be replaced by a consistent estimator.

7.2. Data-Dependent Robust Critical Values: Type 1

Here we improve on the LF critical value by employing an ICS procedure that uses the data to determine whether b is finite. If b is deemed to be finite, i.e., π is weakly identified (or unidentified), then the LF critical value is used. Otherwise, the standard asymptotic critical value is used. This ICS critical value is analogous to the generalized moment selection method used in Andrews and Soares (2010) for moment inequality models.

The ICS procedure chooses between the identification categories $\mathcal{IC}_0 : ||b|| < \infty$ and $\mathcal{IC}_1 : ||b|| = \infty$. The statistic used for identification-category selection is

$$A_n = \left(n \widehat{\beta}_n' \widehat{\Sigma}_{\beta\beta,n}^{-1} \widehat{\beta}_n \right)^{1/2}, \tag{7.4}$$

where $\widehat{\Sigma}_{\beta\beta,n}$ is the upper left $d_\beta \times d_\beta$ block of $\widehat{\Sigma}_n$ and $\widehat{\Sigma}_n$ is the estimator of the covariance matrix defined in (6.1). We use A_n to assess the strength of identification.

Let $\{\kappa_n : n \geq 1\}$ be a sequence of constants, i.e., tuning parameters, that diverges to infinity as $n \to \infty$. One selects $\mathcal{IC}_0$ if $A_n \leq \kappa_n$ and one selects $\mathcal{IC}_1$ otherwise. Under $\mathcal{IC}_0$, A_n is $O_p(1)$. Hence, one consistently selects $\mathcal{IC}_0$ provided κ_n diverges to infinity. We assume:

Assumption K. (i) $\kappa_n \to \infty$ and (ii) $\kappa_n / n^{1/2} \to 0$.

For example, $\kappa_n = (d_\beta \ln n)^{1/2}$, which is analogous to the BIC penalty term, satisfies Assumption K.

Using the ICS procedure described above, the type 1 robust CI with nominal level $1-\alpha$ is obtained by inverting a test based on $\mathcal{T}_n$ with critical value $\widetilde{c}_{\mathcal{T},1-\alpha,n}$ defined by

$$\widetilde{c}_{\mathcal{T},1-\alpha,n} = \begin{cases} c^{LF}_{\mathcal{T},1-\alpha} & \text{if } A_n \leq \kappa_n \\ c_{\mathcal{T},1-\alpha}(\infty) & \text{if } A_n > \kappa_n. \end{cases} \tag{7.5}$$

The type 1 robust critical value $\widetilde{c}_{\mathcal{T},1-\alpha,n}$ can be improved by employing NI and/or plug-in versions of it. They are defined by replacing H by $H(v)$, $\widehat{H}_n$, or $H(v) \cap \widehat{H}_n$, as in Section 7.1. The type 1 NI robust critical value is denoted $\widetilde{c}_{\mathcal{T},1-\alpha,n}(v)$ for $v \in V_r$.

7.3. Data-Dependent Robust Critical Values: Type 2

Next, we consider a type 2 robust critical value that does not require the tuning parameter κ_n to diverge to infinity as $n \to \infty$. In consequence, asymptotic size-correction factors Δ_1 and Δ_2 can be introduced. These size correction factors are designed to improve the asymptotic approximations. The type 2 robust critical value also provides a continuous transition from a weak-identification critical value to a strong-identification critical value using a transition function $s(x)$. This robust critical value is akin to the method employed in Andrews and Jia (2008) for moment inequality models.

Let $s(x)$ be a continuous function on $[0, \infty)$ that satisfies: (i) $0 \leq s(x) \leq 1$, (ii) $s(x)$ is non-increasing in x, (iii) $s(0) = 1$, and (iv) $s(x) \to 0$ as $x \to \infty$. Examples of transition functions include (i) $s(x) = \exp(-c \cdot x)$ for some $c > 0$ and (ii) $s(x) = (1 + c \cdot x)^{-1}$ for some $c > 0$.[27] In the ARMA example, we use the function $s(x) = \exp(-x/2)$.

The type 2 robust critical value is

$$\widehat{c}_{\mathcal{T},1-\alpha,n} = \begin{cases} c^{LF}_{\mathcal{T},1-\alpha} + \Delta_1 & \text{if } A_n \leq \kappa \\ c_{\mathcal{T},1-\alpha}(\infty) + \Delta_2 + [c^{LF}_{\mathcal{T},1-\alpha} + \Delta_1 - c_{\mathcal{T},1-\alpha}(\infty) - \Delta_2] \cdot s(A_n - \kappa) & \text{if } A_n > \kappa, \end{cases} \tag{7.6}$$

where $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$ are defined below. When $A_n \leq \kappa$, $\widehat{c}_{\mathcal{T},1-\alpha,n}$ equals the LF critical value $c^{LF}_{\mathcal{T},1-\alpha}$ plus a size-correction factor Δ_1. When $A_n > \kappa$, $\widehat{c}_{\mathcal{T},1-\alpha,n}$ is a convex combination of $c^{LF}_{\mathcal{T},1-\alpha} + \Delta_1$ and $c_{\mathcal{T},1-\alpha}(\infty) + \Delta_2$, where Δ_2 is another size-correction factor and the weight given to the standard critical value $c_{\mathcal{T},1-\alpha}(\infty)$ increases with the strength of identification, as measured by $A_n - \kappa$.

[27] The asymptotic size results given in Theorem 7.1 below also hold for the abrupt transition function $s(x) = 1 - 1(x > 0)$, which is discontinuous at $x = 0$, provided one adds the assumption that $P(A(h) = \kappa) = 0 \ \forall h \in H$, where $A(h)$ is defined in (7.7) below. The latter condition is satisfied in most examples.

Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$, $A_n \to_d A(h)$, where $A(h)$ is defined by

$$A(h) = \left(\tau_\beta(\pi^*; \gamma_0, b)' \Sigma_{\beta\beta}^{-1}(\pi^*; \gamma_0) \tau_\beta(\pi^*; \gamma_0, b)\right)^{1/2}, \quad (7.7)$$

where π^* abbreviates $\pi^*(\gamma_0, b)$ and $\tau_\beta(\pi; \gamma_0, b)$ and $\Sigma_{\beta\beta}(\pi; \gamma_0)$ are defined in (5.5) and (6.4), respectively.[28,29]

For any Δ_1 and Δ_2, under $\gamma_n \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$, the asymptotic null rejection probability of a test based on the statistic $\mathcal{T}_n$ and the robust critical value $\widehat{c}_{\mathcal{T},1-\alpha,n}$ is shown to equal

$$\begin{aligned} NRP(\Delta_1, \Delta_2; h) &= P(\mathcal{T}(h) > c_B \ \& \ A(h) \le \kappa) + P(\mathcal{T}(h) > c_S(h) \ \& \ A(h) > \kappa) \\ &= P(\mathcal{T}(h) > c_B) + P(\mathcal{T}(h) \in (c_S(h), c_B] \ \& \ A(h) > \kappa), \text{ where} \\ c_B &= c_{\mathcal{T},1-\alpha}^{LF} + \Delta_1, \\ c_S(h) &= c_{\mathcal{T},1-\alpha}(\infty) + \Delta_2 + (c_{\mathcal{T},1-\alpha}^{LF} + \Delta_1 - c_{\mathcal{T},1-\alpha}(\infty) - \Delta_2) \cdot s(A(h) - \kappa), \end{aligned} \quad (7.8)$$

"B" denotes Big, and "S" denotes Small.

The constants Δ_1 and Δ_2 are chosen such that $NRP(\Delta_1, \Delta_2; h) \le \alpha \ \forall h \in H$. In particular, we define

$$\begin{aligned} \Delta_1 &= \sup_{h \in H_1} \Delta_1(h), \text{ where } \Delta_1(h) \ge 0 \text{ solves } NRP(\Delta_1(h), 0; h) = \alpha \\ &\quad \text{or } \Delta_1(h) = 0 \text{ if } NRP(0, 0; h) < \alpha \text{ and} \\ H_1 &= \{(b, \gamma_0) : (b, \gamma_0) \in H \ \& \ ||b|| \le ||b_{\max}|| + D\}, \text{ and} \end{aligned}$$

$$\begin{aligned} \Delta_2 &= \sup_{h \in H} \Delta_2(h), \text{ where } \Delta_2(h) \text{ solves } NRP(\Delta_1, \Delta_2(h); h) = \alpha \\ &\quad \text{or } \Delta_2(h) = 0 \text{ if } NRP(\Delta_1, 0; h) < \alpha. \end{aligned} \quad (7.9)$$

By definition $b_{\max}$ is such that $c_{\mathcal{T},1-\alpha}(h)$ is maximized over $h \in H$ at $h_{\max} = (b_{\max}, \gamma_{\max}) \in$

[28] The convergence in distribution follows from Theorem 5.1(a) and Assumption V1.

[29] In the vector β case, $\Sigma_{\beta\beta}(\pi; \gamma_0)$ is replaced by $\Sigma_{\beta\beta}(\pi, \omega^*(\pi; \gamma_0, b); \gamma_0)$ in (7.7), where $\Sigma_{\beta\beta}(\pi, \omega; \gamma_0)$ is defined in (6.5) and $\omega^*(\pi; \gamma_0, b)$ is defined in (6.9). When the type 2 robust critical value is considered in the vector β case, h is defined to include $\omega_0 \in R^{d_\beta}$ with $||\omega_0|| = 1$ as an element, i.e., $h = (b, \gamma_0, \omega_0)$ and $H = \{h = (b, \gamma_0, \omega_0) : ||b|| < \infty, \gamma_0 \in \Gamma \text{ with } \beta_0 = 0, ||\omega_0|| = 1\}$.

H for some $\gamma_{\max} \in \Gamma$ and D is a positive constant, such as 1.[30,31,32] As defined, Δ_1 and Δ_2 can be computed sequentially, which is computationally convenient.

The adjustment via Δ_1 size corrects for b values that are at or near $b_{\max}$. Size correction is needed here because the ICS statistic A_n is larger than κ with a positive probability asymptotically even under sequences of true parameters for which the LF critical value is needed to achieve correct asymptotic size.

The adjustment via Δ_2 size corrects for relatively large values of b. Size correction may be needed here to handle the difference between the ideal critical value for the given value of b and the robust critical value that is determined by the transition function $s(A_n - \kappa)$. Typically, this discrepancy is small and only a small adjustment Δ_2 is needed.

Given the definitions of Δ_1 and Δ_2, the rejection probability is close to the nominal level α when h is close to $h_{\max}$ (due to the adjustment with Δ_1) and when $||b||$ is large (due to the adjustment with Δ_2).

The type 2 robust critical value defined here can be improved by employing NI and/or plug-in versions of it. The NI robust critical value is defined by replacing H by $H(v)$ (defined in (7.2)) in (7.9) and in the definitions of $h_{\max}$ and $b_{\max}$, which are then denoted $b_{\max}(v)$ and $h_{\max}(v)$. The set H_1 is replaced by $H_1(v) = \{(b, \gamma_0) : (b, \gamma_0) \in H(v)$ & $||b|| \leq \sup_{v_0 \in V_r} ||b_{\max}(v_0)|| + D\}$.[33] The constants Δ_1, Δ_2, $\Delta_1(h)$, and $\Delta_2(h)$ in (7.9) are then denoted $\Delta_1(v)$, $\Delta_2(v)$, $\Delta_1(h, v)$, and $\Delta_2(h, v)$. By definition, for any $v \in V_r$, $NRP(\Delta_1(v), \Delta_2(v); h) \leq \alpha$ for all $h \in H(v)$. The NI robust critical value is denoted $\widehat{c}_{\mathcal{T},1-\alpha,n}(v)$.

For example, consider the construction of a type 2 NI robust CI for the parameter π. For each value of $v \in \Pi$, one first obtains the LF critical value $c^{LF}_{\mathcal{T},1-\alpha}(v)$ and then one calculates $\Delta_1(v)$ and $\Delta_2(v)$ based on $c^{LF}_{\mathcal{T},1-\alpha}(v)$ and the asymptotic distribution of $\mathcal{T}_n$ and A_n under the null $H_0 : \pi_0 = v$.

[30] When $NRP(0, 0; h) > \alpha$, a unique solution $\Delta_1(h)$ typically exists because $NRP(\Delta_1, 0; h)$ is always non-increasing in Δ_1 and is typically strictly decreasing and continuous in Δ_1. If no exact solution to $NRP(\Delta_1(h), 0; h) = \alpha$ exists, then $\Delta_1(h)$ is taken to be any value for which $NRP(\Delta_1(h), 0; h) \leq \alpha$ and $\Delta_1(h) \geq 0$ is as small as possible. Analogous comments apply to the equation $NRP(\Delta_1, \Delta_2(h); h) = \alpha$ and the definition of $\Delta_2(h)$.

[31] When the LF critical value is achieved at $||b|| = \infty$, i.e., $c_{\mathcal{T},1-\alpha}(\infty) \geq \sup_{h \in H} c_{\mathcal{T},1-\alpha}(h)$, the standard asymptotic critical value $c_{\mathcal{T},1-\alpha}(\infty)$ yields a test or CI with correct asymptotic size and constants Δ_1 and Δ_2 are not needed. Hence, here we consider the case where $||b_{\max}|| < \infty$. If $\sup_{h \in H} c_{\mathcal{T},1-\alpha}(h)$ is not attained at any point $h_{\max}$, then $b_{\max}$ can be taken to be any point such that $c_{\mathcal{T},1-\alpha}(h_{\max})$ is arbitrarily close to $\sup_{h \in H} c_{\mathcal{T},1-\alpha}(h)$ for some $h_{\max} = (b_{\max}, \gamma_{\max}) \in H$.

[32] In practice, we find that $D = 1$ works well and the results are not sensitive to the choice of D.

[33] In the definition of $H_1(v)$, the upper bound on $||b||$ does not vary with v, which improves the smoothness of $\Delta_1(v)$ as a function of v.

A plug-in version of the type 2 robust critical value requires the replacement of H by $\widehat{H}_n$ throughout (7.9), where $\widehat{H}_n$ is defined as in Section 7.1. Similarly, a plug-in version of the type 2 NI robust critical value is defined like the type 2 NI robust critical value but with H replaced by $H(v) \cap \widehat{H}_n$ throughout.

Note that for a type 2 NI robust CI or CS for β, under semi-strong or strong identification, $\Delta_1(v) \to 0$ and $\Delta_2(v) \to 0$ as $||b|| \to \infty$, and the NI robust critical value converges to the standard critical value.

For any given value of κ, the type 2 robust CI has correct asymptotic size due to the choice of Δ_1 and Δ_2. In consequence, we choose κ based on the false coverage probabilities (FCP's) of the robust CI. An FCP of a CI for $r(\theta)$ is the probability that the CI includes a value different from the true value $r(\theta)$. Small FCP's are closely linked to short CI's, see Pratt (1961).

The method we use to choose κ is to minimize the average asymptotic FCP of the robust CI at a chosen set of points.[34] We are interested in a robust CI for $r(\theta)$. Let $\mathcal{K}$ denote the set of κ values from which we select. First, for given $h \in H$, we choose a null value $v_{H_0}(h)$ that differs from the true value $v_0 = r(\theta_0)$ (where $h = (b, \gamma_0)$ and $\gamma_0 = (\theta_0, \phi_0)$). The null value $v_{H_0}(h)$ is selected such that the robust CI based on a reasonable choice of κ, such as $\kappa = 1.5$ or 2, has a FCP that is in a range of interest, such as close to 0.50.[35,36] Second, we compute the FCP of the value $v_{H_0}(h)$ for each robust CI with $\kappa \in \mathcal{K}$. Third, we repeat steps one and two for each $h \in \mathcal{H}$, where $\mathcal{H}$ is a representative subset of H.[37] The optimal choice of κ is the value that minimizes over $\mathcal{K}$ the average FCP at $v_{H_0}(h)$ over $h \in \mathcal{H}$.

[34] For t and Wald CI's, asymptotic FCP's follow from the results in this paper and AC3. For QLR CI's, the results currently in AC2 do not cover non-null parameter values. Hence, we compute FCP's for a large, but finite, sample size when determining κ. For example, in the ARMA(1, 1) example, we use $n = 500$.

[35] For reasonable choices, the value of κ used to obtain $v_{H_0}(h)$ typically has very little effect on the final comparison across different values of κ. For example, this is true in the ARMA(1, 1) example considered below.

[36] When b is close to 0, the FCP may be larger than 0.50 for all admissible v due to weak identification. In such cases, $v_{H_0}(h)$ is taken to be the admissible value that minimizes the FCP for the selected value of κ that is being used to obtain $v_{H_0}(h)$.

[37] When $r(\theta) = \pi$ or $r(\theta) = \pi + \beta$, we do not include h values in $\mathcal{H}$ for which $b = 0$ because when $b = 0$ there is no information about π and it is not necessarily desirable to have a small FCP.

7.4. Asymptotic Size of Robust t CI's

In this section, we show that the LF and data-dependent robust CI's defined above have correct asymptotic size when $\mathcal{T}_n$ equals the t statistic or its absolute value. Analogous results for robust QLR and Wald CI's are given in AC2 and AC3, respectively.

The asymptotic size results of this section rely on the following df continuity conditions, which are not restrictive in most examples.

Assumption LF. (i) The df of $\mathcal{T}(h)$ is continuous at $c_{\mathcal{T},1-\alpha}(h)$ $\forall h \in H$.
(ii) If $c^{LF}_{\mathcal{T},1-\alpha} > c_{\mathcal{T},1-\alpha}(\infty)$, $c^{LF}_{\mathcal{T},1-\alpha}$ is attained at some $h_{\max} \in H$.

Assumption NI-LF. (i) The df of $\mathcal{T}(h)$ is continuous at $c_{\mathcal{T},1-\alpha}(h,v)$ $\forall h \in H(v)$, $\forall v \in V_r$.
(ii) For some $v \in V_r$, $c^{LF}_{\mathcal{T},1-\alpha}(v) = c_{\mathcal{T},1-\alpha}(\infty)$ or $c^{LF}_{\mathcal{T},1-\alpha}(v)$ is attained at some $h_{\max} \in H$.

For $h \in H$, define

$$\widehat{c}_{\mathcal{T},1-\alpha}(h) \tag{7.10}$$
$$= \begin{cases} c^{LF}_{\mathcal{T},1-\alpha} + \Delta_1 & \text{if } A(h) \leq \kappa \\ c_{\mathcal{T},1-\alpha}(\infty) + \Delta_2 + [c^{LF}_{\mathcal{T},1-\alpha} + \Delta_1 - c_{\mathcal{T},1-\alpha}(\infty) - \Delta_2] \cdot s(A(h) - \kappa) & \text{if } A(h) > \kappa. \end{cases}$$

Note that $\widehat{c}_{\mathcal{T},1-\alpha}(h)$ equals $\widehat{c}_{\mathcal{T},1-\alpha,n}$ with $A(h)$ in place of A_n. It is shown in the proof of Theorem 7.1 below that the asymptotic distribution of $\widehat{c}_{\mathcal{T},1-\alpha,n}$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $||b|| < \infty$ is the distribution of $\widehat{c}_{\mathcal{T},1-\alpha}(h)$.

Define $\widehat{c}_{\mathcal{T},1-\alpha}(h,v)$ analogously to $\widehat{c}_{\mathcal{T},1-\alpha}(h)$, but with $c^{LF}_{\mathcal{T},1-\alpha}$, Δ_1, and Δ_2 replaced by $c^{LF}_{\mathcal{T},1-\alpha}(v)$, $\Delta_1(v)$, and $\Delta_2(v)$, respectively, for $v \in V_r$. The asymptotic distribution of $\widehat{c}_{\mathcal{T},1-\alpha,n}(v)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $||b|| < \infty$ is the distribution of $\widehat{c}_{\mathcal{T},1-\alpha}(h,v)$.

Assumption Rob2. (i) $P(\mathcal{T}(h) = \widehat{c}_{\mathcal{T},1-\alpha}(h)) = 0$ $\forall h \in H$.
(ii) If $\Delta_2 > 0$, $NRP(\Delta_1, \Delta_2; h^*) = \alpha$ for some point $h^* \in H$, where Δ_1 and Δ_2 are defined in (7.9).

Assumption NI-Rob2. (i) $P(\mathcal{T}(h) = \widehat{c}_{\mathcal{T},1-\alpha}(h,v)) = 0$ $\forall h \in H(v)$, $\forall v \in V_r$.
(ii) For some $v \in V_r$, $\Delta_2(v) = 0$ or $NRP(\Delta_1(v), \Delta_2(v); h^*) = \alpha$ for some point $h^* \in H(v)$, where $\Delta_1(v)$ and $\Delta_2(v)$ are defined after (7.9).

For $\mathcal{T}_n$ equal to the t statistic $|T_n|$, T_n, or $-T_n$, we have $\mathcal{T}(h)$ equals $|T(h)|$, $T(h)$, or $-T(h)$, respectively, the quantile $c_{\mathcal{T},1-\alpha}(h)$ equals $c_{|t|,1-\alpha}(h)$, $c_{t,1-\alpha}(h)$, or $c_{-t,1-\alpha}(h)$ defined just below (6.14), the quantile $c_{\mathcal{T},1-\alpha}(\infty)$ equals $z_{1-\alpha/2}$, $z_{1-\alpha}$, or $z_{1-\alpha}$, and the quantities $c^{LF}_{\mathcal{T},1-\alpha}$, $c^{LF}_{\mathcal{T},1-\alpha}(v)$, $\widetilde{c}_{\mathcal{T},1-\alpha,n}$, $\widetilde{c}_{\mathcal{T},1-\alpha,n}(v)$, $\widehat{c}_{\mathcal{T},1-\alpha,n}$, $\widehat{c}_{\mathcal{T},1-\alpha,n}(v)$, $\widehat{c}_{\mathcal{T},1-\alpha}(h)$, and $\widehat{c}_{\mathcal{T},1-\alpha}(h,v)$ are defined as above with $\mathcal{T} = |t|$, t, or $-t$, respectively.

Theorem 7.1. *Suppose Assumptions* A, B1-B3, C1-C8, D1-D3, R1, *and* V1-V2 *hold and* $d_r = 1$. *Then, the nominal* $1-\alpha$ *symmetric two-sided, upper one-sided, and lower one-sided robust* t *CI's all have* $AsySz = 1-\alpha$ *when based on the following critical values*: (a) LF, (b) NI-LF, (c) *type* 1 *robust,* (d) *type* 1 NI *robust,* (e) *type* 2 *robust, and* (f) *type* 2 NI *robust, provided the following additional Assumptions hold, respectively*: (a) LF, (b) NI-LF, (c) K *and* V3, (d) K *and* V3, (e) Rob2, *and* (f) NI-Rob2, *where* $\mathcal{T}(h)$ *in Assumptions* LF, NI-LF, Rob2, *and* NI-Rob2 *is equal to* $|T(h)|$, $T(h)$, *and* $-T(h)$ *in the two-sided, upper one-sided, and lower-sided cases, respectively.*

Comments. 1. Plug-in versions of the robust CI's considered in Theorem 7.1 also have asymptotically correct size under continuity assumptions on $c_{\mathcal{T},1-\alpha}(h)$ that typically are not restrictive. For brevity, we do not provide formal results here.

2. If part (ii) of Assumption LF, NI-LF, Rob2, or NI-Rob2 does not hold, then the corresponding part of Theorem 7.1 still holds, but with $AsySz \geq 1-\alpha$. For example, Assumptions LF(ii) and Rob2(ii) fail in the unusual case that $c^{LF}_{\mathcal{T},1-\alpha} = \infty$ and Assumptions NI-LF(ii) and NI-Rob2(ii) fail if $c^{LF}_{\mathcal{T},1-\alpha}(v) = \infty \ \forall v \in V_r$.

8. QLR Confidence Sets and Tests

In this section, we introduce CS's based on the quasi-likelihood ratio (QLR) statistic. For brevity, assumptions and theoretical results for the QLR procedures are given in AC2. However, we define QLR procedures here because numerical results are reported for them in the ARMA example section.

We consider CS's for a function $r(\theta)$ $(\in R^{d_r})$ of θ obtained by inverting QLR tests. The function $r(\theta)$ is assumed to be smooth and to be of the form

$$r(\theta) = \begin{bmatrix} r_1(\psi) \\ r_2(\pi) \end{bmatrix}, \tag{8.1}$$

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

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,

$$\widetilde{\theta}_n(v) \in \Theta,\ r(\widetilde{\theta}_n(v)) = v,\ \text{and } Q_n(\widetilde{\theta}_n(v)) = \inf_{\theta \in \Theta : r(\theta) = v} Q_n(\theta) + o(n^{-1}). \quad (8.2)$$

The asymptotic distribution of the restricted estimator $\widetilde{\theta}_n = \widetilde{\theta}_n(v_n)$ for $v_n = r(\theta_n)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ is derived in AC2.

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

$$QLR_n(v) = 2n(Q_n(\widetilde{\theta}_n(v)) - Q_n(\widehat{\theta}_n))/\widehat{s}_n, \quad (8.3)$$

where $\widehat{s}_n$ is a real-valued scaling factor that is employed in some cases to yield a QLR statistic that has an asymptotic $\chi^2_{d_r}$ null distribution under strong identification. See AC2 for details.

Let $c_{n,1-\alpha}(v)$ denote a nominal level $1 - \alpha$ critical value to be used with the QLR test statistic. It may be stochastic or non-stochastic. The usual choice, based on the asymptotic distribution of the QLR statistic under standard regularity conditions, is the $1 - \alpha$ quantile of the $\chi^2_{d_r}$ distribution:

$$c_{n,1-\alpha}(v) = \chi^2_{d_r, 1-\alpha}. \quad (8.4)$$

AC2 determines the asymptotic size of the standard QLR CS.

Critical values that deliver robust QLR CS's for $r(\theta)$ that have correct asymptotic size can be constructed using the approach of Section 7. Details are in AC2.

Given a critical value $c_{n,1-\alpha}(v)$, the nominal level $1 - \alpha$ QLR CS for $r(\theta)$ is

$$CS^{QLR}_{r,n} = \{v \in r(\Theta) : QLR_n(v) \leq c_{n,1-\alpha}(v)\}. \quad (8.5)$$

Example 1 (cont.). We consider tests and CS's involving functions of π and β. In consequence, a key assumption in AC2, Assumption RQ2(ii), holds. This assumption is needed for the QLR statistic to have a $\chi^2_{d_r}$ asymptotic null distribution under strong identification (and is a standard assumption in the literature where strong identification is assumed). It holds in this example because $V(\gamma_0)$ and $J(\gamma_0)$ are block diagonal (after re-ordering their rows and columns) between the (β, π) and ζ parameters and the blocks of $V(\gamma_0)$ and $J(\gamma_0)$ that correspond to the (β, π) parameters are equal, see (4.30) and

(4.32). In consequence, $\widehat{s}_n = 1$ in this example and the standard critical value is $\chi^2_{d_r,1-\alpha}$.

For a test concerning the MA parameter π, by Theorem 5.1(b) of AC2, the asymptotic null distribution of the QLR statistic under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $|b| < \infty$ is the distribution of

$$2\left(\xi(\pi_0;\gamma_0,b) - \inf_{\pi\in\Pi}\xi(\pi;\gamma_0,b)\right) \tag{8.6}$$
$$= -\left(\sum_{j=0}^{\infty}\pi_0^j Z_j - (1-\pi_0^2)^{-1}b\right)^2(1-\pi_0^2) + \inf_{\pi\in\Pi}\left(\sum_{j=0}^{\infty}\pi^j Z_j - (1-\pi_0\pi)^{-1}b\right)^2(1-\pi^2)$$

where $\{Z_j : j \geq 0\}$ are i.i.d. $N(0,1)$ random variables, $\xi(\pi;\gamma_0,b)$ is defined for this example in (4.21), and the equality in (8.6) uses the simplifications in (6.12). This limit distribution only depends on b and π_0.

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, the QLR statistic has a χ^2_1 asymptotic null distribution by Theorem 5.2(b) of AC2 and (5.18) of AC2.

For tests concerning the AR parameter $\pi + \beta$, the QLR statistic has the same asymptotic null distribution as given above for tests concerning the MA parameter π. This holds by Comment 2 to Theorem 5.1 of AC2. $\square$

9. ARMA Example: Simulation Results

In this section, we provide asymptotic and finite-sample simulation results for the ARMA(1, 1) model.

The model is given in (3.2) with $\varepsilon_t \sim N(0,1)$. The optimization parameter spaces for the MA and AR parameters are $[-.85, .85]$ and $[-.90, .90]$, respectively. The true parameter spaces are $[-.80, .80]$ and $[-.85, .85]$, respectively. These choices are designed to cover a broad range of parameters, but to avoid unit root and boundary effects. The parameter spaces satisfy Assumptions B1 and B2.

The sample sizes considered in the simulation include n = 100, 250, and 500. The number of simulation repetitions used for both the asymptotic and finite-sample simulations is $50,000$. For brevity, details concerning the computations and some of the results are provided in Appendix D of AC1-SM.

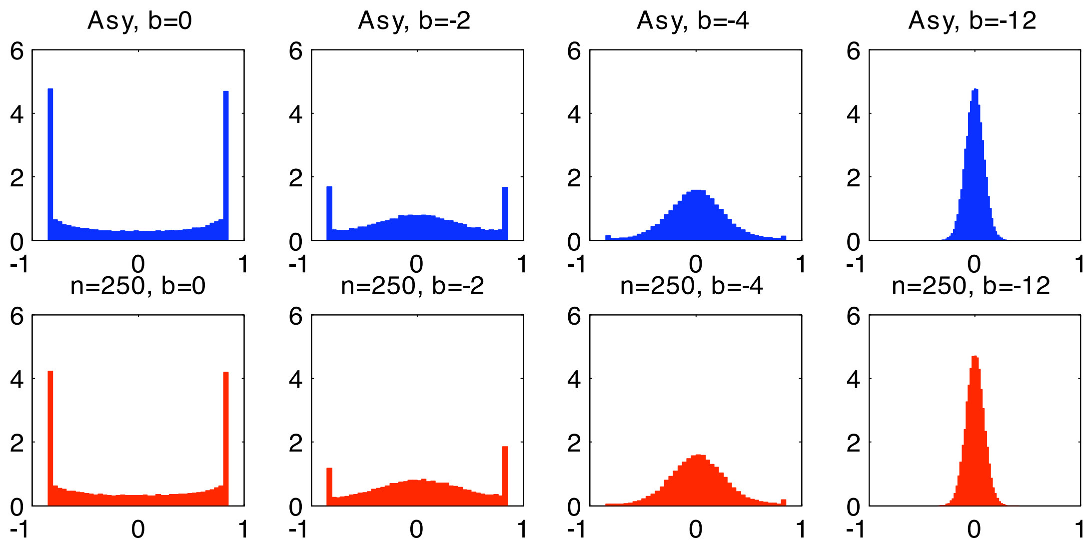

Figure 1. Asymptotic and Finite-Sample (n=250) Densities of the Estimator of the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0$.

9.1. Estimators

Figures 1 and 2 provide the asymptotic and finite-sample densities of the ML estimator of the MA parameter π when the true π value, π_0, is 0.0 and 0.4, respectively. Each Figure gives the densities for $b = 0, -2, -4$, and -12, where b indexes the magnitude of the difference β between the AR and MA parameters.[38] Specifically, for the finite-sample results, $b = n^{1/2}\beta$. In these Figures, the finite-sample size considered is $n = 250$. Note that for $n = 250$, the values $b = 0, -2, -4$, and -12 correspond to β being 0.0, -0.13, -0.25, and -0.76, respectively. For $n = 100$, these b values correspond to β being 0.0, -0.2, -0.4, and -1.2, respectively. Figure S-1 of AC1-SM provides analogous results for $\pi_0 = 0.7$.

Figures 1 and 2 show that the ML estimator has a distribution that is very far from a normal distribution in the unidentified and weakly-identified cases. In these cases, there is a build-up of mass at the boundaries of the optimization space. There also is a bias towards 0 when $\pi_0 > 0$. This is most pronounced when $\pi_0 = 0.7$.

Figures 1 and 2 indicate that the asymptotic approximations developed here work strikingly well. There are some differences, but they are relatively small.

Figure 3 provides analogous results to those of Figure 2 for the ML estimator of β,

[38]The asymptotic densities in Figures 1 and 2 are invariant to the sign of b.

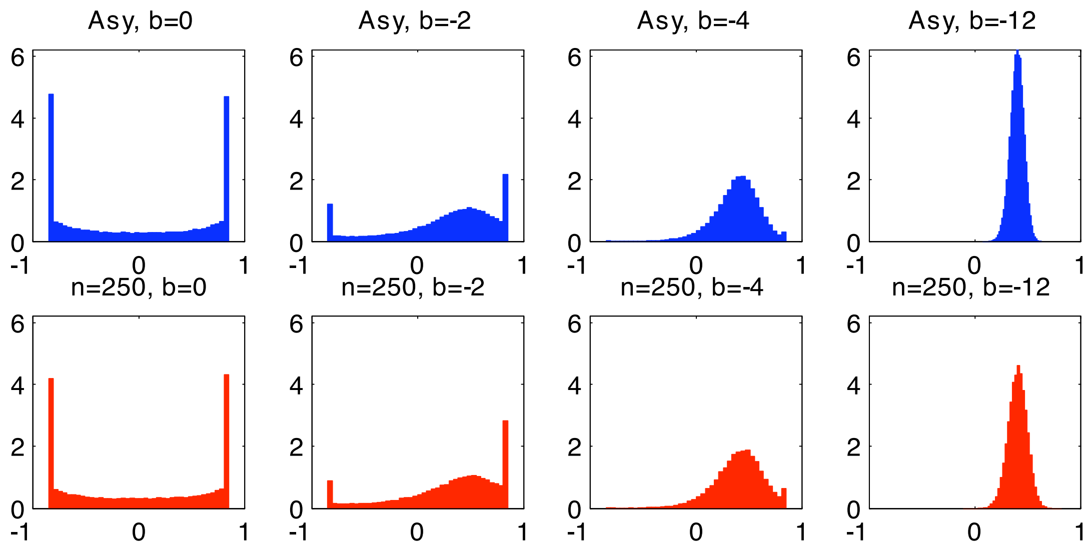

Figure 2. Asymptotic and Finite-Sample (n=250) Densities of the Estimator of the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.4$.

the difference between the AR and MA parameters. In Figure 3, $\pi_0 = 0.4$. Figure 3 shows a very pronounced bi-modal distribution in the unidentified case and a side-lobe in one weakly-identified case. Again, the asymptotic approximations are found to work exceptionally well.

Analogous results for the ML estimator of the AR parameter are provided in Figures S-9 to S-11 in AC1-SM. The asymptotic distributions are identical to those for the MA parameter.[39] The differences between the asymptotic and finite-sample results are larger for the AR parameter than the MA, mainly at the boundary points with $b = 0$, but they are still quite close.

In sum, the estimation results demonstrate a substantial effect on the distributions of the parameter estimators due to lack of identification or weak identification. The asymptotic theory developed in the paper does a very good job in capturing these effects.

[39] Under weak identification, this is because the estimator of π is not consistent, whereas the estimator of β is $n^{1/2}$-consistent. In consequence, the asymptotic distribution of the estimator of $\rho = \pi + \beta$ is the same as that of π.

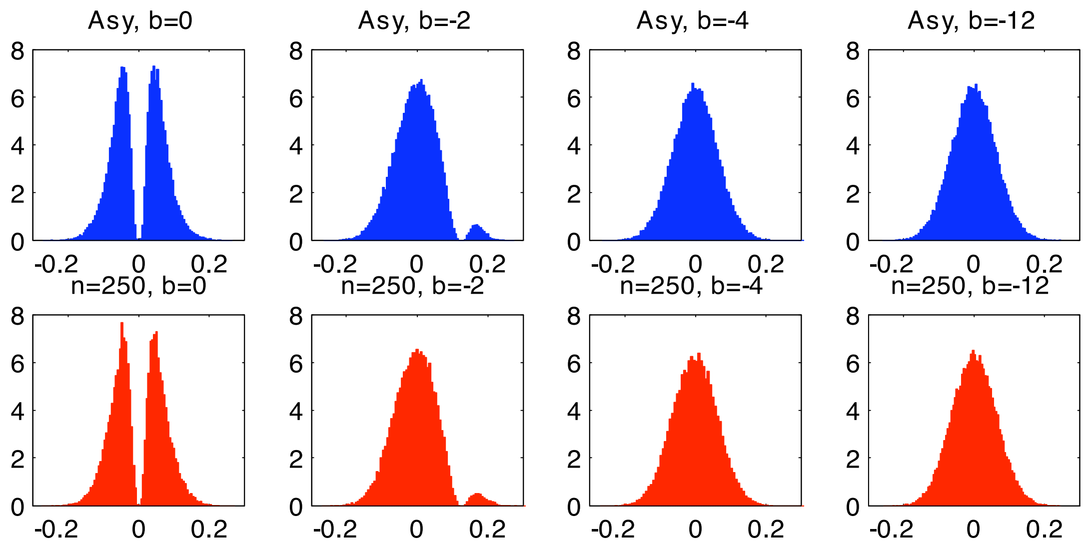

Figure 3. Asymptotic and Finite-Sample (n=250) Densities of the Estimator of β (Centered at the True Value) in the ARMA(1, 1) Model when $\pi_0 = 0.4$.

9.2. Test Statistics and Standard CI's

Figures 4 and 5 provide the asymptotic and finite-sample ($n = 250$) densities of the t and QLR statistics, respectively, for tests concerning the MA parameter π for $\pi_0 = 0.4$ and $b = 0$, -2, -4, and -12. The black lines in Figures 4 and 5 are the standard normal and χ^2_1 densities, respectively, which are the strong-identification asymptotic densities of the test statistics.

Figure 4 shows that the t statistic has a noticeably non-normal shape due to skewness and kurtosis for small $|b|$, although it is much less non-normal than the distribution of the corresponding estimator.[40] Figure 5 indicates that the QLR statistic is well approximated by a χ^2_1 distribution even under weak identification. This suggests that the QLR statistic yields tests and CI's that are substantially less sensitive to weak identification than t-based tests and CI's are.

Figures analogous to Figures 4 and 5, but for $\pi_0 = 0.0$ and 0.7, are given in Figures S2-S5 of AC1-SM. Similar patterns emerge, although the skewness of the $|t|$ statistic varies with π_0. Figures analogous to these, but for the t and QLR statistics for ρ, rather than π, are given in Figures S-12 to S-17 of AC1-SM. The Figures for ρ are quite similar

[40] The distributions of the estimator of π and the t statistic for π are not the same up to a scale shift even asymptotically. This occurs because the variance estimator that appears in the t statistic involves an estimator of π, which is not consistent when $|b| < \infty$. It is random even in the limit.

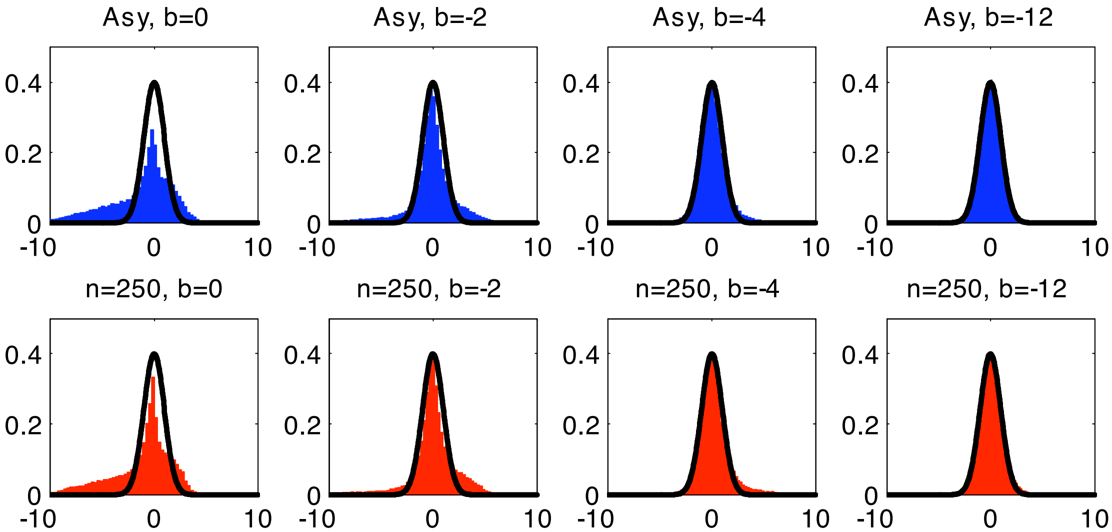

Figure 4. Asymptotic and Finite-Sample (n=250) Densities of the t Statistic for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.4$ and the Standard Normal Density (Black Line).

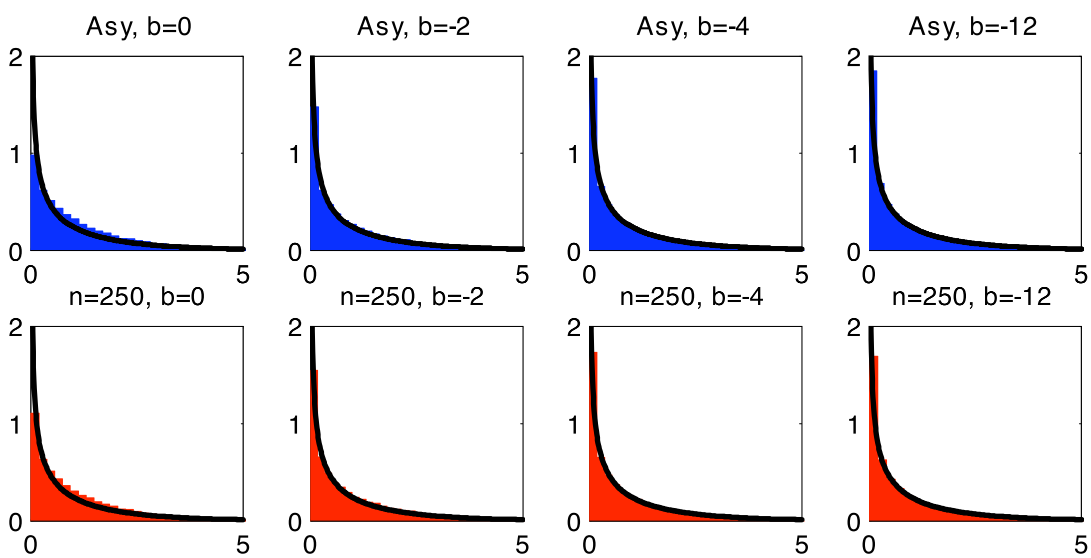

Figure 5. Asymptotic and Finite-Sample (n=250) Densities of the QLR Statistic for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.4$ and the χ_1^2 Density (Black Line).

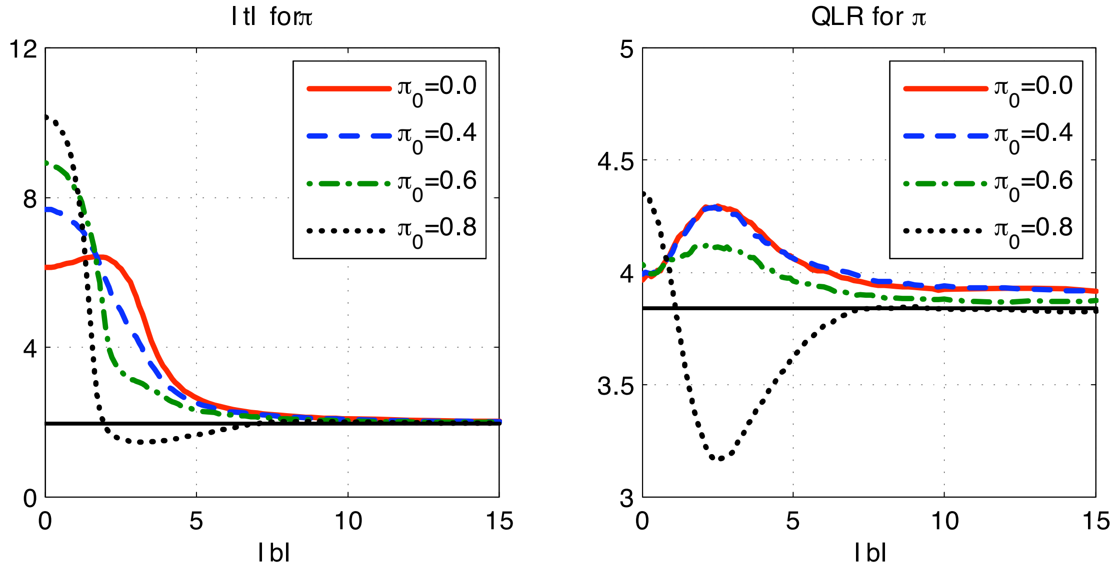

Figure 6. Asymptotic 0.95 Quantiles of the $|t|$ and QLR Statistics for Tests Concerning the MA Parameter π in the ARMA(1, 1) Model.

to those for π.

Figure 6 provides graphs of the 0.95 asymptotic quantiles of the $|t|$ and QLR statistics concerning the MA parameter π as a function of $|b|$ for $\pi_0 = 0.0$, 0.4, 0.6, and 0.8.[41] For both statistics, for small to medium $|b|$ values, the graphs exceed the 0.95 quantiles under strong identification (given by the horizontal black lines). This implies that tests and CI's that employ the standard critical values (based on the normal or χ^2_1 distribution) have incorrect size. For the t statistic, however, the exceedance is very large and much larger than for the QLR statistic. For example, for $\pi_0 = 0.8$ and $b = 0$, the quantile is roughly 10, whereas for strong identification ($|b| = \infty$) it is roughly 2. In contrast, for the QLR statistic, for $\pi_0 = 0.8$ and $b = 0$, the quantile is roughly 4.4, whereas for strong identification it is roughly 3.8.

The asymptotic quantiles given in Figure 6 also apply to the t and QLR statistics concerning the AR parameter $\rho = \pi + \beta$. Hence, no additional quantile graphs are provided for ρ.

Figure 7 provides asymptotic quantile graphs for t and QLR statistics concerning the parameter β that are analogous to those in Figure 6 for π. Again we find that for small to medium values of $|b|$ the graphs exceed the 0.95 strong-identification quantile.

[41] The asymptotic quantiles are invariant to the sign of b, but the finite-sample quantiles are not.

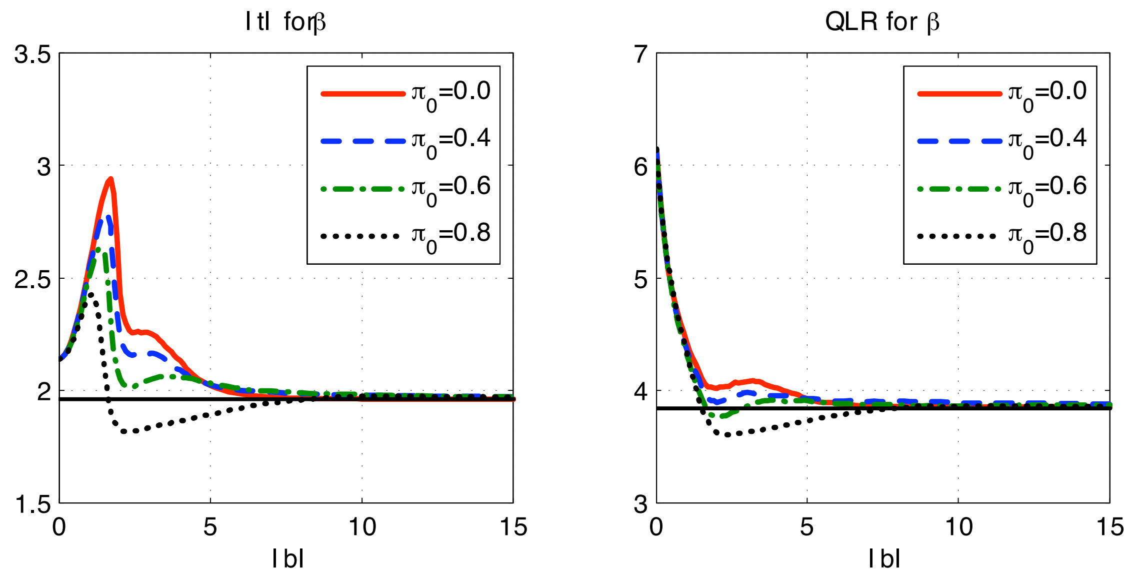

Figure 7. Asymptotic 0.95 Quantiles of the $|t|$ and QLR Statistics for Tests Concerning β in the ARMA(1, 1) Model.

For tests concerning β, however, the t and QLR graphs are much more similar to each other than for those concerning π.

Figure 8 reports asymptotic and finite-sample CP's of nominal 95% standard $|t|$ and QLR CI's (which employ normal and χ_1^2 critical values, respectively) for the MA parameter π. The CP's are given as a function of b (≤ 0), for true $\pi_0 = 0.0$, for $n = 100$, 250, 500, and ∞ (i.e., asymptotic).[42] As one would predict given Figure 6, the CP's of the $|t|$ CI are very low for $|b|$ values less than 10. For $b = 0$, the asymptotic and finite-sample CP's are all below 0.60. Hence, the size of this nominal 95% CI is less than 0.60 asymptotically and in finite samples. More specifically, Table II provides the minimum over b asymptotic CP's for a range of true π_0 values. It shows that the asymptotic size of the $|t|$ CI for π is 0.523.[43] Note that the asymptotic CP's in Figure 8 provide a very good approximation to the finite-sample CP's.

Figure 8 and Table II show that the under-coverage of the standard QLR CI for π is much less severe than for the $|t|$ CI. Table II shows that the asymptotic size of the

[42]In Figures 8-10, the graphs for $n = 100$ are not given for all values of b because b is restricted by the parameter space. The same is true for the graphs for $n = 250$ in Figures 8 and 9. See Appendix D of AC1-SM for details. These parameter space restrictions are responsible for the wiggles that occur in some of the $n = 100$ and 250 graphs in Figures 8-10 near the right end of the graphs.

[43]This is based on a grid of π_0 values that is the same as the grid of 21 π_{H_0} values given in Table V.

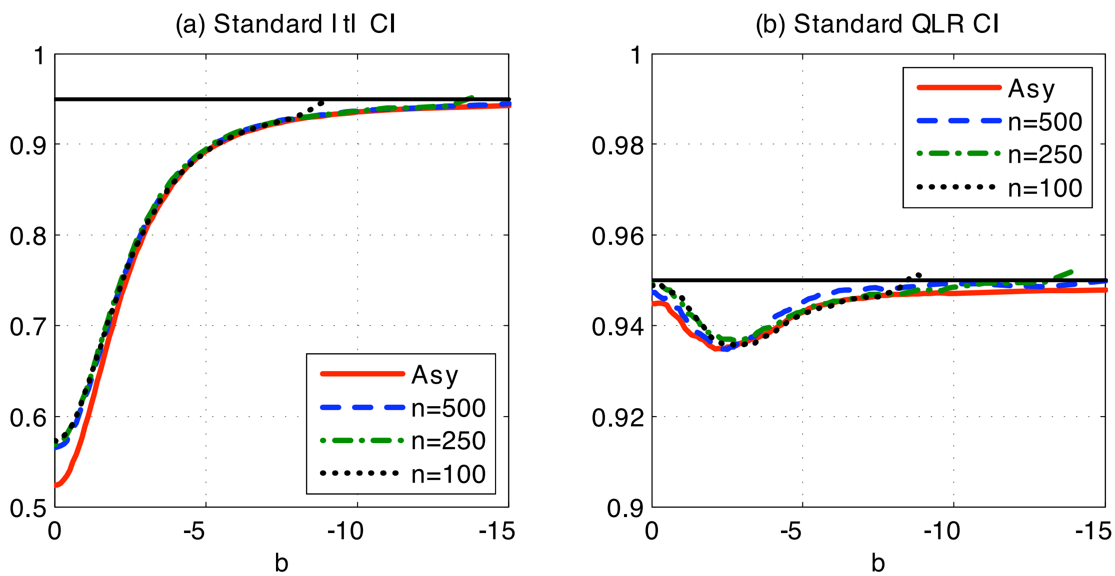

Figure 8. Coverage Probabilities of Standard $|t|$ and QLR CI's for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0$.

Table II. Asymptotic Coverage Probabilities (Minimum over b) of Nominal 95% Standard CI's for π and ρ in the ARMA(1, 1) Model

π_0/ρ_0	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	Asy Size
$\|t\|$	0.523	0.527	0.534	0.552	0.578	0.612	0.642	0.643	0.627	0.523
QLR	0.935	0.933	0.933	0.934	0.935	0.936	0.940	0.941	0.933	0.933

nominal 95% standard QLR CI for π is 0.933. Figure 8 shows that the asymptotic and finite-sample CP's are very close for all b when $\pi_0 = 0$.

Figures S-6 and S-7 in AC1-SM provide analogous graphs to those in Figure 8 for $\pi_0 = 0.4$ and 0.7, rather than $\pi_0 = 0.0$. The graphs are all similar in shape.

For the AR parameter, analogous figures and tables to those described above for the MA parameter are provided in Figures S-18 to S-20 in AC1-SM. Generally speaking, the results for the AR parameter are quite similar to those for the MA parameter. Indeed, the asymptotic results are identical. Hence, Table II applies to standard CI's for both π and ρ.

9.3. Robust Confidence Intervals

Next, we consider CI's that are robust to weak identification. We focus on the type 2 NI robust CI's, which are defined in Section 7.3. For comparative purposes, we also provide some results for NI-LF CI's. For the type 2 NI robust CI's, we employ the transition function $s(x) = \exp(-x/2)$ and the constants $\kappa = 1.5$ and $D = 1$. The choices of $s(x)$ and D were determined via some experimentation to be good choices in terms of yielding CP's that are relatively close to the nominal size 0.95 across different values of b. Given $s(x)$ and D, the choice of κ was determined using the method described at the end of Section 7.3 based on minimizing average FCP's. The details are given in Appendix D of AC1-SM. It turns out that a wide range of κ values yields similar average FCP's, see Tables S-II, S-III, S-V, and S-VI in Appendix D of AC1-SM, so the particular choice of $\kappa = 1.5$ is not at all crucial.[44,45]

Figures 9 and 10 report the asymptotic and finite-sample CP's of the type 2 NI robust $|t|$ and QLR CI's for the MA parameter π as a function of b (≤ 0) for $\pi_0 = 0.0$ and 0.4, respectively. Figure 9 shows that the CP's of both the $|t|$ and QLR CI's are greater than or equal to 0.95 for all b when $\pi_0 = 0.0$. However, the QLR CI is closer to being similar, both asymptotically and in finite-samples. Only for $|b| \leq 3$ are its CP's greater than 0.95. The asymptotic approximations perform very well in Figure 9.

For the QLR CI, the results in Figure 10 for $\pi_0 = 0.4$ are quite similar to those in Figure 9. The CP's of the robust QLR CI are greater than or equal to 0.95 for all b and they exceed 0.95 only for $|b| \leq 4$. The asymptotic and finite-sample CP's are very close. For the $|t|$ CI, however, there is a greater discrepancy between the asymptotic and finite-sample results than when $\pi_0 = 0.0$. In addition, there is some under-coverage. For $n = 100$, the CP's of $|t|$ CI are as low as 0.93 for some b values. However, the magnitude of the under-coverage of the robust $|t|$ CI is very small compared to that of the standard $|t|$ CI.

Figure S-8 of AC1-SM provides results analogous to those of Figures 9 and 10, but for $\pi_0 = 0.7$. The results for the robust QLR test are similar to those in Figures 9 and 10. But, for the robust $|t|$ CI, there are larger differences between the asymptotic and finite-sample CP's and there is greater finite-sample under-coverage.

Figures S-21 to S-23 in AC1-SM report results analogous to those of Figures 9, 10, and S-8 but for robust CI's for ρ, rather than π. The results for the robust QLR CI's

[44] The reason is that if κ is changed, the constants τ_1 and τ_2 change in a manner that substantially offsets the effect of the change in κ. This occurs because, for any given κ, the constants τ_1 and τ_2 must yield a CI with the desired size.

[45] The value $\kappa = 1.5$ is used for all CI's considered, whether they are $|t|$ or QLR-based and whether they are for π or ρ. This value of κ minimizes the average FCP measured to two significant digits for all cases considered, see the tables in Appendix D of AC1-SM.

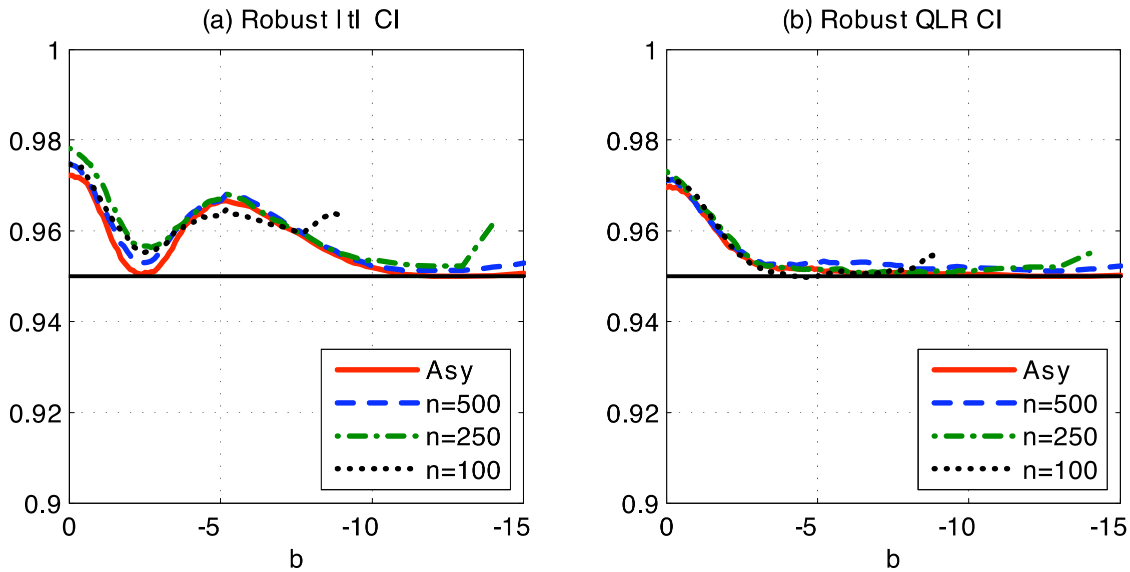

Figure 9. Coverage Probabilities of Robust $|t|$ and QLR CI's for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

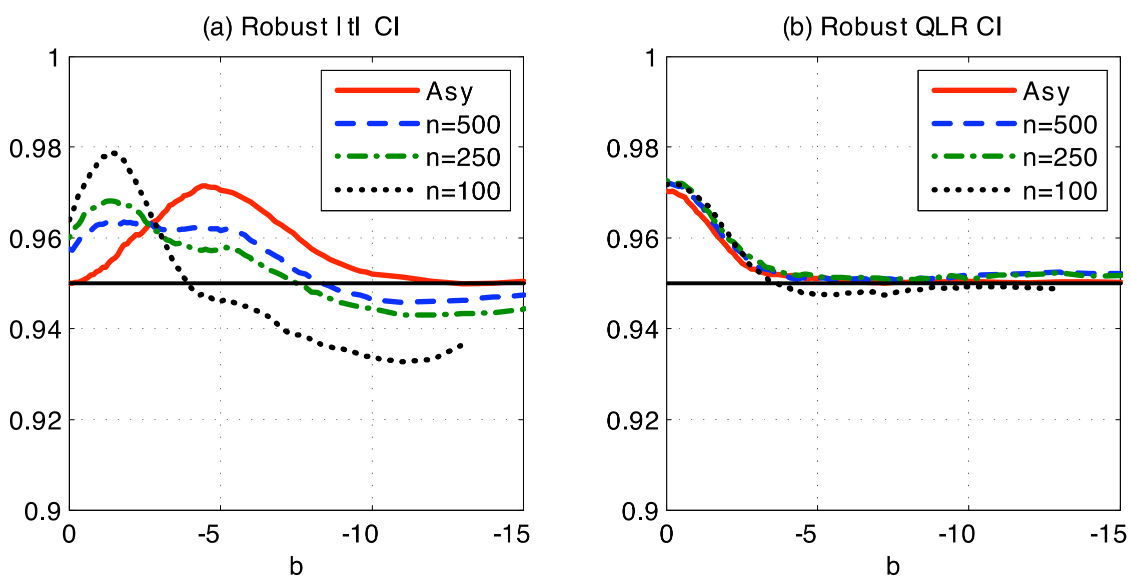

Figure 10. Coverage Probabilities of Robust $|t|$ and QLR CI's for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.4$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

Table III. Finite-Sample Coverage Probabilities (Minimum over b) of Nominal 95% CI's for π and ρ in the ARMA(1, 1) Model, $n = 250$

		$\|t\|$			QLR		
		Std	LF	Rob	Std	LF	Rob
MA	$\pi_0 = 0.0$	0.569	0.965	0.952	0.937	0.951	0.951
	$\pi_0 = 0.4$	0.613	0.961	0.943	0.937	0.953	0.951
	$\pi_0 = 0.7$	0.673	0.962	0.930	0.944	0.953	0.946
AR	$\rho_0 = 0.0$	0.573	0.967	0.955	0.937	0.952	0.950
	$\rho_0 = 0.4$	0.632	0.966	0.953	0.939	0.954	0.953
	$\rho_0 = 0.8$	0.660	0.965	0.952	0.936	0.954	0.950

for ρ are quite similar to those for π. The results for the robust $|t|$ CI's for ρ show larger differences. They exhibit no under-coverage, but the regions of b values where the CP's exceed 0.95 are larger.

Table III provides a summary of the finite-sample ($n = 250$) CP's of the CI's based on critical values that are standard (normal or χ^2_1), NI-LF, and type 2 NI robust. It provides results for $|t|$ and QLR CI's for both π and ρ. The standard $|t|$ CI's under-cover considerably. The standard QLR CI's only under-cover by a small amount. The NI-LF $|t|$ CI's over-cover by a small amount. The type 2 NI robust $|t|$ CI's are close to 0.95 except for some under-coverage for π when $\pi_0 = 0.4$ and 0.7. The NI-LF and type 2 NI robust QLR CI's are quite close to 0.95.

Table S-I of AC1-SM provides analogous results to Table III, but for $n = 100$ and 500. The results for the standard CI's are very similar to those in Table III. The discrepancies between the CP's and 0.95 for the NI-LF and type 2 NI robust $|t|$ CI's are magnified for $n = 100$ and lessened for $n = 500$. The CP's for the NI-LF and type 2 NI robust QLR CI's are quite close to 0.95 for $n = 100$ and 500.

Table IV provides finite-sample FCP results for the NI-LF and type 2 NI robust CI's for the MA parameter π for $n = 500$. The true values considered are $\pi_0 = 0.0$, 0.4, and 0.7 and $b = -2, -5, -10$, and $-\infty$. The null values π_{H_0} are provided in the Table. They are selected so that the robust QLR CI has FCP close to 0.50 for those cases where that is possible. (When $b = 0$ or $|b|$ is small, all CI's have FCP greater than 0.50 for all values of π_{H_0} in the parameter space.) Table IV shows that the $|t|$ statistic combined with the NI-LF critical value yields a CI whose FCP's are very high—close to 1.0 for most values of b and π_0. This illustrates the poor performance of LF critical values when a substantial amount of size correction is required. The NI-LF critical value performs much better in terms of FCP's when combined with the QLR statistic (because much

Table IV. Finite-Sample False Coverage Probabilities of 95% Least Favorable and Robust $|t|$ and QLR CI's for the MA parameter π in the ARMA(1, 1) Model, $n = 500$

	$\pi_0 = 0.0$				$\pi_0 = 0.4$				$\pi_0 = 0.7$				
b	-2	-5	-10	$-\infty$	-2	-5	-10	$-\infty$	-2	-5	-10	$-\infty$	Avg
π_{H_0}	0.800	0.410	0.200	0.048	0.000	0.010	0.205	0.290	0.000	0.460	0.570	0.615	
$\|t\|$													
LF	0.97	1.00	1.00	1.00	0.93	0.96	1.00	1.00	0.76	0.99	1.00	1.00	0.97
Rob	0.95	0.78	0.56	0.90	0.91	0.64	0.49	0.49	0.68	0.57	0.44	0.44	0.65
QLR													
LF	0.68	0.51	0.55	0.52	0.88	0.52	0.55	0.55	0.59	0.53	0.54	0.53	0.58
Rob	0.67	0.50	0.51	0.49	0.89	0.50	0.51	0.51	0.62	0.51	0.51	0.51	0.56

less size-correction is needed). The type 2 NI robust critical values work quite well in terms of FCP's with both the $|t|$ and QLR statistics. Overall, the type 2 NI robust QLR CI performs best, followed closely by the NI-LF QLR CI, followed by the type 2 NI robust $|t|$ CI.

Analogous results to those in Table IV, but for the AR parameter ρ, are provided in Table S-IV of AC1-SM. Most of the results are quite similar.

Table V. Values of NI LF Critical Values and $\Delta_1(\pi_{H_0})$ and $\Delta_2(\pi_{H_0})$ for Size Correction in the ARMA(1, 1) Model

$\|t\|$	π_{H_0}/ρ_{H_0}	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	$c^{LF}_{\|t\|,.95}(\pi_{H_0})$	6.43	6.43	6.43	6.43	6.57	6.81	7.09	7.39	7.69	8.01	8.31
	$\Delta_1(\pi_{H_0})$	1.22	1.21	1.19	1.12	0.90	0.64	0.32	0.22	0.20	0.19	0.20
	$\Delta_2(\pi_{H_0})$	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.06	0.05	0.06	0.06
	π_{H_0}/ρ_{H_0}	0.55	0.60	0.625	0.65	0.675	0.70	0.725	0.75	0.775	0.80	0.825
	$c^{LF}_{\|t\|,.95}(\pi_{H_0})$	8.62	8.94	9.09	9.24	9.40	9.55	9.70	9.86	10.01	10.17	10.25
	$\Delta_1(\pi_{H_0})$	0.21	0.22	0.22	0.23	0.24	0.25	0.25	0.26	0.26	0.27	0.26
	$\Delta_2(\pi_{H_0})$	0.05	0.03	0.02	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.01
QLR	π_{H_0}/ρ_{H_0}	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	$c^{LF}_{QLR,.95}(\pi_{H_0})$	4.30	4.31	4.32	4.32	4.33	4.32	4.31	4.30	4.29	4.28	4.25
	$\Delta_1(\pi_{H_0})$	0.60	0.62	0.71	0.73	0.76	0.81	0.82	0.77	0.68	0.64	0.55
	$\Delta_2(\pi_{H_0})$	0.08	0.08	0.08	0.09	0.10	0.10	0.08	0.09	0.09	0.09	0.09
	π_{H_0}/ρ_{H_0}	0.55	0.60	0.625	0.65	0.675	0.70	0.725	0.75	0.775	0.80	0.825
	$c^{LF}_{QLR,.95}(\pi_{H_0})$	4.21	4.13	4.08	4.07	4.09	4.12	4.16	4.22	4.29	4.36	4.37
	$\Delta_1(\pi_{H_0})$	0.57	0.55	0.54	0.45	0.29	0.18	0.07	0.09	0.11	0.12	0.12
	$\Delta_2(\pi_{H_0})$	0.06	0.04	0.04	0.03	0.04	0.04	0.04	0.02	0.01	0.00	0.00

Table V provides the $c^{LF}_{\mathcal{T},1-\alpha}(v)$, $\Delta_1(v)$, and $\Delta_2(v)$ values necessary to compute the type 2 NI robust critical values for the $|t|$ and QLR test statistics for computing CI's for the MA and AR parameters. (The same values apply to both the MA and AR parameters.) In this case, v denotes the null hypothesis value of π (or ρ), which we denote by π_{H_0} (or ρ_{H_0}) in the Table. For π_{H_0} (or ρ_{H_0}) values between those given in Table V, linear interpolation can be used.

References

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

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.

——— (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. McFadden. Amsterdam: North-Holland.

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

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

Andrews, D. W. K. and X. Cheng (2007): "Supplemental Material for 'Estimation and Inference with Weak, Semi-strong, and Strong Identification,' " unpublished manuscript, Cowles Foundation, Yale University.

——— (2008a): "Maximum Likelihood Estimation and Inference with Weak, Semi-strong, and Strong Identification," manuscript in preparation, Cowles Foundation, Yale University.

——— (2008b): "GMM Estimation and Inference with Weak, Semi-strong, and Strong Identification," manuscript in preparation, Cowles Foundation, Yale University.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2009): "Generic Results for Establishing the Asymptotic Size of Confidence Intervals and Tests," unpublished manuscript, Cowles Foundation, Yale University.

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

Andrews, D. W. K. and P. Jia (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," Cowles Foundation Discussion Paper No. 1676, Yale University.

Andrews, D. W. K., X. Liu, and W. Ploberger (1998): "Tests for White Noise Against Alternatives with Both Seasonal and Non-seasonal Serial Correlation," *Biometrika,* 85, 727–740.

Andrews, D. W. K. and C. J. McDermott (1995): "Nonlinear Econometric Models with Deterministically Trending Variables," *Review of Economic Studies,* 62, 343–360.

Andrews, D. W. K. and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica,* 62, 1383–1414.

——— (1996): "Testing for Serial Correlation Against an ARMA(1, 1) Process," *Journal of the American Statistical Association,* 91, 1331-1342.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica,* 78, 119-157.

Andrews, D. W. K. and J. H. Stock (2007): "Inference with Weak Instruments," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society,* Vol. III, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge, UK: Cambridge University Press.

Ansley, C. F. and P. Newbold (1980): "Finite Sample Properties of Estimators for Autoregressive Moving Average Models," *Journal of Econometrics,* 13, 159-183.

Antoine, B. and E. Renault (2009): "Efficient GMM with Nearly-Weak Instruments," *Journal of Econometrics,* 12, S135-S171.

——— (2010): "Efficient Inference with Poor Instruments, a General Framework," ed. by D. Giles and A. Ullah, *Handbook of Empirical Economics and Finance.* Oxford, UK: Taylor and Francis, forthcoming.

Bai, J. (1997): "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics,* 79, 551-563.

Chan, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *Annals of Statistics,* 21, 520-533.

Chan, K. S. and R. S. Tsay (1998): "Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model," *Biometrika,* 85, 413-426.

Cheng, X. (2008): "Robust Confidence Intervals in Nonlinear Regression under Weak Identification," unpublished working paper, Department of Economics, Yale University.

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

Choi, I. and P. C. B. Phillips (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics,* 51, 113-150.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative," *Biometrika,* 64, 247-254.

——— (1987): "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternatives," *Biometrika,* 74, 33-43.

Dong, Y. (2009): "Endogenous Regressor Binary Choice Models without Instruments, with an Application to Migration," unpublished manuscript, Department of Economics, California State University at Fullerton.

Drton, M. (2009): "Likelihood Ratio Tests and Singularities," *Annals of Statistics,* 37, 979-1012.

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

Elliott, G. and U. K. Müller (2007): "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics,* 141, 1196-1218.

——— (2008): "Pre and Post Break Parameter Inference," unpublished manuscript, Department of Economics, Princeton University.

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

Ghysels, E., A. Sinko, and R. Valkanov (2007): "MIDAS Regressions: Further Results and New Directions," *Econometric Reviews,* 26, 53-90.

Hahn, J. and G. Kuersteiner (2002): "Discontinuities of the Weak Instruments Limiting Distribution," *Economic Letters,* 75, 325-331.

Han, S. (2009): "Identification and Inference in a Bivariate Probit Model with Weak Instruments," unpublished manuscript, Department of Economics, Yale University.

Hannan, E. J. (1982): "Testing for Autocorrelation and Akaike's Criterion," in *Essays in Statistical Science,* ed. by J. Gani and E. J. Hannan. *Journal of Applied Probability,* Special Volume 19A.

Hansen, B. E. (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica,* 64, 413-430.

——— (2000): "Sample Splitting and Threshold Estimation," *Econometrica,* 68, 575-603.

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

——— (2005): "Testing Parameters in GMM without Assuming That They Are Identified," *Econometrica,* 73, 1103-1123.

Kleibergen, F. and S. Mavroeidis (2009): "Inference on Subsets of Parameters in GMM without Assuming Identification," unpublished manuscript, Department of Economics, Brown University.

Liu, X. and Y. Shao (2003): "Asymptotics for Likelihood Ratio Tests Under Loss of Identifiability," *Annals of Statistics,* 31, 807-832.

Ma, J. and C. R. Nelson (2006): "Valid Inference for a Class of Models Where Standard Inference Performs Poorly; Including Nonlinear Regression, ARMA, GARCH, and Unobserved Components," unpublished manuscript, Department of Economics, U. of Washington.

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

Nelson, C. R. and R. Startz (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica,* 58, 967-976.

——— (2007): "The Zero-information-limit Condition and Spurious Inference in Weakly Identified Models," *Journal of Econometrics,* 138, 47-62.

Nelson, F. and L. Olson (1978): "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables," *International Economic Review,* 19, 695-709.

Newey, W. K. (1987): "Efficient Estimation of Limited Dependent Variable Models with Endogenous Regressors," *Journal of Econometrics,* 36, 231-250.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics,* Vol. IV, Ch. 36, ed. by R. F. Engle and D. McFadden. Amsterdam: North-Holland.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

Park, J. Y. and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory,* 4, 468-497.

——— (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory,* 15, 269-298.

——— (2001): "Nonlinear Regressions with Integrated Time Series," *Econometrica,* 69, 117-161.

Phillips, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory,* 5, 181-240.

Picard, D. (1985): "Testing and Estimating Change-point in Time Series," *Advances in Applied Probability,* 17, 841-867.

Pollard, D. (1984): *Convergence of Stochastic Processes*. New York: Springer–Verlag.

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

Pratt, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association,* 56, 541-567.

Rivers, D. and Q. H. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics,* 39, 347-366.

Sargan, J. D. (1983): "Identification and Lack of Identification," *Econometrica,* 51, 1605-1633.

Shi, X. and P. C. B. Phillips (2009): "Nonlinear Cointegrating Regression Under Weak Identification," unpublished manuscript, Cowles Foundation, Yale University.

Smith, R. J. and R. W. Blundell (1986): "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica,* 54, 679-685.

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica,* 65, 557-586.

Stock, J. H. and J. H. Wright (2000): "GMM with Weak Instruments," *Econometrica,* 68, 1055-1096.

Tripathi, G. T. (1999): "A Matrix Extension of the Cauchy-Schwarz Inequality," *Economics Letters,* 63, 1-3.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes.* New York: Springer.