Estimation and Inference with Weak Identification

Donald W. K. Andrews*
Cowles Foundation
Yale University

Xu Cheng
Department of Economics
University of Pennsylvania

First Draft: August, 2007
Revised: March 4, 2010

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0751517.

Abstract

This paper analyses 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 a method 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), and minimum distance (MD) 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 LS estimator of a nonlinear regression model, a LS estimator of a smooth transition threshold autoregressive model, the instrumental variables estimator of a nonlinear regression model with endogeneity, and the ML estimator of an ARMA (1, 1) model.

Keywords: Asymptotic size, confidence set, estimator, identification, nonlinear models, 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 $\beta h(X_i, \pi) + Z_i'\zeta$.

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.

In addition to the nonlinear regression model, other examples that are considered in the paper include a smooth transition threshold autoregressive (STAR) model, which includes smooth transition switching regression models, the ARMA (1, 1) model, and a nonlinear regression model with endogenous regressors. Han (2009) shows that, via reparametrization, a simple bivariate probit model with endogeneity falls into the class of models considered here. Other examples include 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 $(t^\pi - 1)/\pi$.[1]

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 (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, 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. 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 and Wald tests, and t and Wald CS's under lack of identification, weak identification, semi-strong identification, and strong identification, as discussed below.

[1] 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).

We investigate the large sample biases of extremum estimators under weak identification. We determine the asymptotic size of standard 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.[2]

We introduce a method of making standard tests and CS's robust to lack of identification, i.e., to have correct asymptotic size. The method is closely related to a method suggested in Andrews (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 idea is to use a testing/model 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.

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 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 in-

[2] 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.

terest display discontinuities in their pointwise asymptotic distributions, see Andrews 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.

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.

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, $\lvert\lvert\beta_n\rvert\rvert = O(n^{-1/2})$)	Weakly identified
II	$\beta_n \to 0$ and $n^{1/2}\lvert\lvert\beta_n\rvert\rvert \to \infty$	Semi-strongly identified
III	$\beta_n \to \beta_0 \neq 0$	Strongly identified

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.

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 (1996a). In contrast, the hypotheses considered in this paper are of a more general type. Here we consider a full range of nonlinear hypotheses concerning (β, ζ, π)—only special cases are of the type just described. 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 $\beta = 0$. Furthermore, even if interest is focussed on β, to obtain asymptotic size results for CS's for β, 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 different from those considered in the literature referred to above.

The weak instrumental variable (IV) literature, e.g., see Nelson and Startz (1990), Dufour (1997), Staiger and Stock (1997), Stock and Wright (2000), Kleibergen (2002, 2005), Moreira (2003), and other papers referenced in Andrews and Stock (2007), is related to the present paper. This is especially true of Staiger and Stock (1997). In the weak IV literature, the criterion functions considered do not depend on the parameters that are the source of weak identification. In the present paper, the criterion functions do. 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 and yields results that cover both the structural and reduced-form parameters, 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).

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).

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.

Phillips (1989) and Choi and Phillips (2001) 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. 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).

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 **??** 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 **??** below),
(vi) the (generalized) first derivative of this quadratic expansion has an asymptotic normal distribution after a matrix rescaling when β is local to the non-identification value 0, and
(vii) the (generalized) Hessian of this quadratic expansion is nonsingular asymptotically after a matrix rescaling 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 $n^{1/2}(\psi_n - \psi_{0,n}) \to (b, 0)$ 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): $||\beta_n||^{-1}(Q_n^c(\pi) - Q_n(\psi_{0,n}, \pi)) \to_p \eta(\pi)$ 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 $\widehat{\pi}_n \to_d \pi^* = \arg\min_{\pi \in \Pi} \xi(\pi)$. 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 $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$: $n^{1/2}(\widehat{\psi}_n - \psi_n) \to_d \tau(\pi^*)$. 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.

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 (**??**) 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 the Wald and 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 Wald and t test for Category I-III sequences.

The analysis of Wald tests for multiple restrictions requires the use of a parameter dependent matrix rotation to separate the effects of randomness in $\widehat{\psi}_n$ and $\widehat{\pi}_n$ because these two estimators have different rates of convergence for sequences $\{\gamma_n\}$ in Categories I and II.

We show that for some multiple nonlinear restrictions the Wald test statistic diverges to infinity in probability under the null hypothesis. Obviously this is not a desirable property and leads to the standard Wald test (for any nominal level $\alpha > 0$) having asymptotic size equal to one.

For Category I sequences, the asymptotic distribution of the t statistic for a scalar 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 scalar 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 Wald and t tests and Wald and t 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 data-dependent critical values that yield Wald and t tests and Wald and t CS'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.

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 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.

The paper focuses on inference when θ is not identified 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 (by the criterion function $Q_n(\theta)$) 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 Assumption A 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. This example is a nonlinear regression model estimated by LS. We use it as a running example to illustrate the more general results. The model is

$$Y_i = \beta \cdot h(X_i, \pi) + Z_i'\zeta + U_i \text{ for } i = 1, ..., n, \tag{3.2}$$

where $h(X_i, \pi) \in R$ is known up to the finite-dimensional parameter $\pi \in R^{d_\pi}$. When the true value of β is 0, (3.2) becomes a linear model and π is not identified.

Suppose the support of X_i is contained in a set $\mathcal{X}$. We assume here that $h(x, \pi)$ is twice continuously differentiable wrt π, $\forall \pi \in \Pi$, $\forall x \in \mathcal{X}$, although the general theory of the paper allows for non-smooth functions. Let $h_\pi(x, \pi) \in R^{d_\pi}$ and $h_{\pi\pi}(x, \pi) \in R^{d_\pi \times d_\pi}$ denote the first-order and second-order partial derivatives of $h(x, \pi)$ wrt π.

The LS sample criterion function is

$$Q_n(\theta) = n^{-1} \sum_{i=1}^{n} U_i^2(\theta)/2, \text{ where } U_i(\theta) = Y_i - \beta h(X_i, \pi) - Z_i'\zeta. \tag{3.3}$$

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

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 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 $i, m \geq 1$.[3,4] 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 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:[5]

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

We assume that the interior of Θ includes the true parameter space Θ^* (see Assumption B1 below). This ensures that the asymptotic distribution of $\widehat{\theta}_n$ is not effected 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

[3]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.

[4]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.

[5]The $o(n^{-1})$ term in (3.5), and in (**??**) and (**??**) 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 (**??**). 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.

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

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

where $T_n(v)$ is a test 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 of a CS for $r(\theta)$ is

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

where $P_\gamma(\cdot)$ denotes the 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 defined as

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

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

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

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 $T_n(\theta)$ under sequences of true parameters $\gamma_n = (\theta_n, \phi_n)$ 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:

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

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 $b \in R^{d_\beta}$, $\{\beta_n\}$ 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 $\{\gamma_n\} \in \Gamma(\gamma_0)$" means "when the true parameters are $\{\gamma_n\} \in \Gamma(\gamma_0)$ for any $\gamma_0 \in \Gamma$;" "under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$" means "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 $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$" means "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 $||\omega_0|| = 1$."

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. (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory*, 4, 458-467.

——— (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.

——— (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.

——— (1997): "Estimation When a Parameter Is on a Boundary: Part II," unpublished manuscript, Cowles Foundation, Yale University.

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

——— (2000): "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space," *Econometrica,* 68, 399–405.

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

——— (2002): "Generalized Method of Moments Estimation When a Parameter Is on a Boundary," *Journal of Business and Economic Statistics*, 20, 530-544.

Andrews, D. W. K., 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 (2009): "Hybrid and Size-Corrected Subsampling Methods," *Econometrica,* 77, 721-762.

——— (2010): "Asymptotic Size and a Problem with Subsampling and with the m Out of n Bootstrap," *Econometric Theory,* 26, forthcoming.

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.

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.

Arcones, M. A. and B. Yu (1994): "Central Limit Theorems for Empirical and U-processes of Stationary Mixing Sequences," *Journal of Theoretical Probability*, 7, 47–71.

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.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford: Oxford University Press.

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.

de Jong, R. M. (1997): "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory*, 13, 353-367.

Doukhan, P., P. Massart, and E. Rio (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annals of the Institute of Henri Poincaré*, 31, 393–427.

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.

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

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

——— (1996b): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays," *Econometric Theory*, 12, 347-359.

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

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimation," *Econometrica*, 50, 1029-1054.

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.

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.

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 (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.

——— (1985): “New Ways to Prove Central Limit Theorems,” *Econometric Theory*, 1, 295-313.

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

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

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.