Journal of Econometrics 173 (2013) 36–56

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

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

Maximum likelihood estimation and uniform inference with sporadic identification failure

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

[a] Cowles Foundation, Yale University, United States

[b] Department of Economics, University of Pennsylvania, United States

ARTICLE INFO

Article history:
Received 17 October 2011
Received in revised form
20 September 2012
Accepted 9 October 2012
Available online 17 October 2012

JEL classification:
C12
C15

Keywords:
Asymptotic size
Binary choice
Confidence set
Estimator
Identification
Likelihood
Nonlinear models
Test
Smooth transition threshold autoregression
Weak identification

ABSTRACT

This paper analyzes the properties of a class of estimators, tests, and confidence sets (CSs) when the parameters are not identified in parts of the parameter space. Specifically, we consider estimator criterion functions that are sample averages and are smooth functions of a parameter θ. This includes log likelihood, quasi-log likelihood, and least squares criterion functions.

We determine the asymptotic distributions of estimators under lack of identification and under weak, semi-strong, and strong identification. We determine the asymptotic size (in a uniform sense) of standard t and quasi-likelihood ratio (QLR) tests and CSs. We provide methods of constructing QLR tests and CSs that are robust to the strength of identification.

The results are applied to two examples: a nonlinear binary choice model and the smooth transition threshold autoregressive (STAR) model.

1. Introduction

This paper provides a set of maximum likelihood (ML) regularity conditions under which the asymptotic properties of ML estimators and corresponding t and QLR tests and confidence sets (CSs) are obtained. The novel feature of the conditions is that they allow the information matrix to be singular in parts of the parameter space. In consequence, the parameter vector is unidentified and weakly identified in some parts of the parameter space, while it is semi-strongly and strongly identified in other parts. The conditions maintain the usual assumption that the log likelihood satisfies a stochastic quadratic expansion. The results also apply to quasi-log likelihood and nonlinear least squares procedures.

Compared to standard asymptotic results in the literature for ML estimators, tests, and CSs, the results given here cover both fixed and drifting sequences of true parameters. The latter are necessary to treat cases of weak identification and semi-strong identification. In particular, they are necessary to determine the asymptotic sizes of tests and CSs (in a uniform sense).

This paper is a sequel to Andrews and Cheng (2012a) (AC1). The method of establishing the results outlined above and in the Abstract is to provide a set of sufficient conditions for the high-level conditions of AC1 for estimators, tests, and CSs that are based on smooth sample-average criterion functions. The high-level conditions in AC1 involve the behavior of the estimator criterion function under certain drifting sequences of distributions. In contrast, the assumptions given here are much more primitive. They only involve mixing, smoothness, and moment conditions, plus conditions on the parameter space.

The paper considers 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. For examples, the nonlinear binary choice model is of the form $Y_i = 1\,(Y_i^* > 0)$ and $Y_i^* = \beta \cdot h(X_i, \pi) + Z_i'\zeta - U_i$, where (Y_i, X_i, Z_i) is observed and $h(\cdot, \cdot)$ is a known function. The STAR model is of the form $Y_t = \zeta_1 + \zeta_2 Y_{t-1} +$

* Corresponding author. Tel.: +1 203 432 3698; fax: +1 203 432 6167.
E-mail address: donald.andrews@yale.edu (D.W.K. Andrews).

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2012.10.003

$\beta \cdot m(Y_{t-1}, \pi) + U_t$, where Y_t is observed and $m(\cdot, \cdot)$ is a known function.

In general, the parameters β, ζ, and π may be scalars or vectors. We determine the asymptotic properties of ML estimators, tests, and CSs under drifting sequences of parameters/distributions. Suppose that 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 ML estimators and test statistics depends on the magnitude of $\|\beta_n\|$. The asymptotic behavior of these statistics varies across three categories of sequences $\{\beta_n : n \geq 1\}$. Category I(a): $\beta_n = 0 \, \forall n \geq 1$, π is unidentified; Category I(b): $\beta_n \neq 0$ and $n^{1/2}\beta_n \to b \in R^{d_\beta}$, π is weakly identified. Category II: $\beta_n \to 0$ and $n^{1/2}\|\beta_n\| \to \infty$, π is semi-strongly identified. Category III: $\beta_n \to \beta_0 \neq 0$, π is strongly identified.

For Category I sequences, we obtain the following results. The estimator of π is inconsistent, the estimator of $\psi = (\beta, \zeta)$ and the t and QLR test statistics have non-standard asymptotic distributions, and the standard tests and CSs (that employ standard normal or χ^2 critical values) have asymptotic null rejection probabilities and coverage probabilities that may or may not be correct, depending on the model.[1] (In many cases, they are not correct.) For Category II sequences, estimators and standard tests and CSs are found to have standard asymptotic properties, but the rate of convergence of the estimator of π is less than $n^{1/2}$. Specifically, the estimators are asymptotically normal and the test statistics have asymptotic chi-squared distributions. For Category III sequences, the estimators and standard tests and CSs have standard asymptotic properties and the estimators converge at rate $n^{1/2}$.

We also consider t and QLR tests and CSs that are robust to the strength of identification. These procedures use different critical values from the standard ones. First, we consider critical values based on asymptotically least favorable sequences of distributions. Next, we consider data-dependent critical values that employ an identification-category selection procedure that determines whether β is near the value 0 that yields lack of identification of π, and if it is, the critical value is adjusted (in a smooth way) to take account of the lack of identification or weak identification. We show that the robust procedures have correct asymptotic size (in a uniform sense). The data-dependent robust critical values yield more powerful tests than the least favorable critical values.

In the numerical results for the STAR and nonlinear binary choice models, π is taken to be a scalar to ease computation. In the STAR model, the transition parameter is fixed, as in the empirical work in Lundbergh and Teräsvirta (2006), and the unknown parameter π is the threshold parameter. The numerical results in both models are summarized as follows. The asymptotic distributions of the estimators of β and π are far from the normal distribution under weak identification and lack of identification. The asymptotic distributions range from being strongly bimodal, to being close to uniform, to being extremely peaked. The asymptotics provide remarkably accurate approximations to the finite-sample distributions.

In the STAR model, the standard t and QLR confidence intervals (CIs) for β have substantial asymptotic size distortions with asymptotic sizes equaling .56 and .72, respectively, for nominal .95 CIs. This is also true for the t and QLR CIs for π, where the asymptotic sizes are .40 and .84, respectively. Note that the size distortions are noticeably larger for the standard t CI than for the QLR CI. In the binary choice model, the standard t and QLR CIs for β have incorrect asymptotic sizes: .68 versus .92, respectively, for nominal .95 CIs. However, the standard t and QLR CIs for π have small and no size distortion, respectively. In both models, the asymptotic sizes provide very good approximations to the finite-sample sizes for the cases considered.

In both models, the robust CIs have correct asymptotic sizes and finite-sample sizes that are quite close to the asymptotic size for the QLR CIs and fairly close for the t CIs. (As mentioned above, for the STAR model, these results are for the case of a fixed transition parameter.)

In sum, the numerical results indicate that the asymptotic results of the paper are quite useful in determining the finite-sample behavior of estimators and standard tests and CIs under weak identification and lack of identification. They are also quite useful in designing robust tests and CIs whose finite-sample size is close to their nominal size.

The results of this paper apply when the criterion function satisfies a stochastic quadratic expansion in the parameter θ. This rules out a number of interesting models that exhibit lack of identification in parts of the parameter space, including regime-switching models, mixture models, abrupt transition structural change models, and abrupt transition threshold autoregressive models, such as in Hansen (2000).[2]

Now, we briefly discuss the literature related to this paper. See AC1 for a more detailed discussion. The following are companion papers to this one: AC1, Andrews and Cheng (2012b) (AC1-SM), and Andrews and Cheng (2011a) (AC3). These papers provide related complementary results to the present paper. AC1 provides results under high-level conditions and analyzes the ARMA(1,1) model in detail. AC1-SM provides proofs for AC1 and related results. AC3 provides results for estimators and tests based on generalized method of moments (GMM) criterion functions. It provides applications to an endogenous nonlinear regression model and an endogenous binary choice model.

Cheng (2008) provides results for a nonlinear regression model with multiple sources of weak identification, whereas the present paper only considers a single source. However, the present paper applies to a much broader range of models.

Tests of $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ are tests in which a nuisance parameter π only appears under the alternative. Such tests have been considered in the literature starting from Davies (1977). The results of this paper cover tests of this sort, as well as tests for a whole range of linear and nonlinear hypotheses that involve (β, ζ, π) and corresponding CSs.

The weak instrument (IV) literature is closely related to this paper. However, papers in that literature focus on criterion functions that are indexed by parameters that do not determine the strength of identification. In contrast, in this paper, the parameter β, which determines the strength of identification of π, appears as one of parameters in the criterion function. Selected papers from the weak IV literature include Nelson and Startz (1990), Dufour (1997), Staiger and Stock (1997), Stock and Wright (2000), Kleibergen (2002, 2005) and Moreira (2003).

Andrews and Mikusheva (2011) and Qu (2011) consider Lagrange multiplier (LM) tests in a maximum likelihood context where identification may fail, with emphasis on dynamic stochastic general equilibrium models. The results of the present paper apply to t and QLR statistics, but not to LM statistics. The consideration of LM statistics is in progress. Andrews and Mikusheva (2012) consider Anderson–Rubin-type tests based on minimum distance statistics for models with weak identification.

Antoine and Renault (2009, 2010) and Caner (2010) consider GMM estimation with IVs that lie in the semi-strong category, using our terminology. Nelson and Startz (2007) and Ma and Nelson (2008) analyze models like those considered in this paper.

[1] Here, by "correct" we mean α or less for tests and $1-\alpha$ or greater for CSs, where α and $1-\alpha$ are the nominal sizes of the tests or CSs.

[2] See AC1 for other references concerning results for these models.