

Inference when a nuisance parameter is weakly identified under the null hypothesis

Stanislav Anatolyev*
New Economic School, Moscow

Abstract

When a nuisance parameter is weakly identified under the null hypothesis, the usual asymptotic theory breaks down and standard tests may exhibit significant size distortions. We provide asymptotic approximations under a drifting parameter DGP for distributions of classical tests and of those designed for the case of complete non-identification. Simulations with a simple SETAR model show that the usual asymptotic theory does fail, although actual sizes of the classical Likelihood Ratio test display surprising robustness to the degree of identification.

JEL classification codes: C12, C13, C22

Key words: weak identification, size distortion, drifting parameter DGP, threshold model.

*Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, Moscow, 117418 Russia.
E-mail: sanatoly@nes.ru. Phone: +7-095-129-1700. Fax: +7-095-129-3722.

1 Introduction and setup

In the standard scenario of testing econometric models, a researcher applies classical asymptotic tests and uses critical values provided by the normal and chi-squared distributions. When there is a non-identified parameter under the null hypothesis, however, the classical tests yield misleading results. In such cases the suitable approach is sharply different: one applies instead non-standard tests developed by Davies (1987), Andrews and Ploberger (1994) and Hansen (1996), among others.

Even in the standard scenario, however, some parameters may not be precisely estimated when the restrictions being tested are imposed, and the situation is essentially close to that characterized by the presence of a non-identified parameter. It arises, for example, when testing for equality of slopes in different regimes in some regime switching model when the corresponding intercepts are close to each other, or the other way round. Even though formally in such circumstances the standard tests are appropriate, their application for typically available samples may lead to drastic size distortions. In this paper, we provide alternative asymptotic approximations to some test statistics under a drifting parameter DGP. This approach has been recently used in the context of linear models with weak instruments (e.g., Staiger and Stock (1997)) and nonlinear models estimated by GMM (e.g., Stock and Wright (2000)). We then run simulations with a simple SETAR model and show that the standard asymptotic theory does fail to approximate well the critical values for most existing tests.

Consider a nonlinear regression model similar to that in Hansen (1996):

$$y_t = x_t' \alpha_0 + h_1(z_t, \gamma_0)' \theta_0 + h_2(z_t, \gamma_0)' \varphi_0 + \varepsilon_t,$$

where α is $k_x \times 1$, θ is $k_1 \times 1$, φ is $k_2 \times 1$, $\gamma \in \Gamma$ is scalar, Γ is a compact set, and zero subscripts refer to true values. The error ε_t is an MDS relative to information embedding present and past x_t and z_t and its own history, and has constant (for simplicity) conditional variance σ^2 . All variables are strictly stationary. The null hypothesis of interest is

$$H_0 : \quad \varphi_0 = 0.$$

Suppose that under H_0 the parameter γ_0 is only weakly identified, which we model by a drifting parameter DGP with $\alpha_0 = \text{const}$ and $\theta_0 = \pi/\sqrt{T}$, where $\pi \neq 0$. When $\pi = 0$, there is complete non-identification under H_0 .

As a leading example, we consider a two regime self-exciting threshold first-order autoregression, and test the equality of AR coefficients in the two regimes. In this case

$$x_t = (1 \ y_{t-1})', \quad (h_1(z_t, \gamma) \ h_2(z_t, \gamma))' = x_t \mathbb{I}[y_{t-1} > \gamma], \quad \alpha = (\mu \ \rho)', \quad \theta = \Delta\mu, \quad \varphi = \Delta\rho,$$

where $\mathbb{I}[\cdot]$ denotes the indicator function, γ is a threshold value, μ and ρ are an intercept and AR coefficient in the first regime, and $\Delta\mu$ and $\Delta\rho$ are their increments across the regimes. The null hypothesis corresponds to testing whether the persistence is the same across the regimes, while the mean is allowed to differ. When the mean shift in the data is small, the threshold parameter γ_0 is weakly identified. Note that the regression function is non-differentiable with respect to γ .

Let vector e and matrices X , $H_1(\gamma)$, $H_{x1}(\gamma)$, $H_2(\gamma)$ and $H(\gamma)$ consist of rows ε_t , x_t' , $h_1(z_t, \gamma)'$, $(x_t', h_1(z_t, \gamma))'$, $h_2(z_t, \gamma)'$ and $h_t(\gamma)' \equiv (x_t', h_1(z_t, \gamma)', h_2(z_t, \gamma))'$, respectively, for $t = 1, \dots, T$. Let us also introduce selector matrices $J_x = (I_{k_x} \ 0_{k_x \times (k_1 + k_2)})'$, $J_1 = (0_{k_1 \times k_x} \ I_{k_1} \ 0_{k_1 \times k_2})'$, $J_2 = (0_{k_2 \times (k_x + k_1)} \ I_{k_2})'$, $J_{12} = (0_{(k_1 + k_2) \times k_x} \ I_{k_1 + k_2})'$, $J_{x1} = (I_{k_x + k_1} \ 0_{(k_x + k_1) \times k_2})'$, and $J_{x2} = \begin{pmatrix} I_{k_x} & & 0_{k_x \times k_2} \\ & 0_{(k_x + k_2) \times k_1} & \\ & & I_{k_2} \end{pmatrix}'$, so that $X = H(\gamma)J_x$, $H_1(\gamma) = H(\gamma)J_1$, etc.

Under suitable conditions spelled out in Hansen (1996),

$$\frac{H(\gamma_1)'H(\gamma_2)}{T} \xrightarrow{AS} Q(\gamma_1, \gamma_2) \equiv E[h_t(\gamma_1)h_t(\gamma_2)']$$

uniformly over $\gamma_1, \gamma_2 \in \Gamma$, and

$$\frac{H(\gamma)'e}{\sqrt{T}} \Rightarrow \Psi(\gamma),$$

where $\Psi(\gamma)$ is a $(k_x + k_1 + k_2)$ -variate Gaussian process with covariance kernel $\sigma^2 Q(\gamma_1, \gamma_2)$, and \Rightarrow denotes weak convergence with respect to the uniform metric.

The model is estimated by minimizing the average squared residual (ASR) via the concentration method: the ASR is minimized for fixed values of γ , and the resulting ASR function is minimized with respect to γ . When γ is fixed, the ASR under H_0 equals

$$\begin{aligned} ASR(\gamma) &= T^{-1} \left(e + X\alpha_0 + H_1(\gamma_0)\pi/\sqrt{T} \right)' \left(I - H(\gamma) \left(H(\gamma)'H(\gamma) \right)^{-1} H(\gamma)' \right) \\ &\quad \times \left(e + X\alpha_0 + H_1(\gamma_0)\pi/\sqrt{T} \right) \\ &= \frac{e'e}{T} - \frac{1}{T} \frac{e'H(\gamma)}{\sqrt{T}} \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'e}{\sqrt{T}} + \frac{1}{T} \pi' \frac{H_1(\gamma_0)'H_1(\gamma_0)}{T} \pi \\ &\quad - \frac{1}{T} \pi' \frac{H_1(\gamma_0)'H(\gamma)}{T} \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'H_1(\gamma_0)}{T} \pi \\ &\quad + \frac{2}{T} \frac{e'H_1(\gamma_0)}{\sqrt{T}} \pi - \frac{2}{T} \frac{e'H(\gamma)}{\sqrt{T}} \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'H_1(\gamma_0)}{T} \pi. \end{aligned}$$

Suppose we impose $\varphi = 0$. Then for fixed γ_C , the ASR under H_0 equals, similarly,

$$\begin{aligned} ASR_C(\gamma_C) &= \frac{e'e}{T} - \frac{1}{T} \frac{e'H_{x1}(\gamma_C)}{\sqrt{T}} \left(\frac{H_{x1}(\gamma_C)'H_{x1}(\gamma_C)}{T} \right)^{-1} \frac{H_{x1}(\gamma_C)'e}{\sqrt{T}} + \frac{1}{T} \pi' \frac{H_1(\gamma_0)'H_1(\gamma_0)}{T} \pi \\ &\quad - \frac{1}{T} \pi' \frac{H_1(\gamma_0)'H_{x1}(\gamma_C)}{T} \left(\frac{H_{x1}(\gamma_C)'H_{x1}(\gamma_C)}{T} \right)^{-1} \frac{H_{x1}(\gamma_C)'H_1(\gamma_0)}{T} \pi \\ &\quad + \frac{2}{T} \frac{e'H_1(\gamma_0)}{\sqrt{T}} \pi - \frac{2}{T} \frac{e'H_{x1}(\gamma_C)}{\sqrt{T}} \left(\frac{H_{x1}(\gamma_C)'H_{x1}(\gamma_C)}{T} \right)^{-1} \frac{H_{x1}(\gamma_C)'H_1(\gamma_0)}{T} \pi. \end{aligned}$$

2 Estimators and their asymptotic behavior

When γ is fixed, the unconstrained ASR under H_0 is

$$\begin{aligned} ASR(\gamma) &= \text{Terms that do not depend on } \gamma \\ &\quad - \frac{1}{T} \left(\frac{H(\gamma)'e}{\sqrt{T}} + \frac{H(\gamma)'H_1(\gamma_0)}{T} \pi \right)' \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \left(\frac{H(\gamma)'e}{\sqrt{T}} + \frac{H(\gamma)'H_1(\gamma_0)}{T} \pi \right). \end{aligned}$$

Thus the minimizer $\hat{\gamma}$ of $ASR(\gamma)$ converges almost surely to

$$\gamma^* = \arg \sup_{\gamma \in \Gamma} \{ (\Psi(\gamma) + Q(\gamma, \gamma_0)J_1\pi)' Q(\gamma, \gamma)^{-1} (\Psi(\gamma) + Q(\gamma, \gamma_0)J_1\pi) \}.$$

Note that if not for the first term in the round brackets, the maximizer would be γ_0 because

$$Q(\gamma_0, \gamma)Q(\gamma, \gamma)^{-1}Q(\gamma, \gamma_0) \leq Q(\gamma_0, \gamma_0)$$

by the matrix Cauchy–Schwarz inequality (e.g., Tripathi (1999)). This fact reflects consistency of $\hat{\gamma}$ when γ_0 is well identified, in which case π is proportional to \sqrt{T} and the first term asymptotically disappears. Under weak identification, the first term in the round brackets of the objective function makes the solution γ^* different from γ_0 and random. Under complete non-identification when $\pi = 0$, the true γ_0 does not enter into the objective function at all.

Analogously, the minimizer $\hat{\gamma}_C$ of $ASR_C(\gamma)$ converges almost surely to

$$\gamma_C^* = \arg \sup_{\gamma \in \Gamma} \left\{ (\Psi(\gamma) + Q(\gamma, \gamma_0)J_1\pi)' J_{x1} (J_{x1}'Q(\gamma, \gamma)J_{x1})^{-1} J_{x1}' (\Psi(\gamma) + Q(\gamma, \gamma_0)J_1\pi) \right\}.$$

Note that $ASR(\hat{\gamma})$ and $ASR_C(\hat{\gamma}_C)$ are consistent for σ^2 under H_0 in spite of the inconsistency of $\hat{\gamma}$ and $\hat{\gamma}_C$.

Next, for fixed γ , under H_0

$$\begin{pmatrix} \hat{\alpha}(\gamma) \\ \hat{\theta}(\gamma) \\ \hat{\varphi}(\gamma) \end{pmatrix} = (H(\gamma)'H(\gamma))^{-1} H(\gamma)' \left(e + X\alpha_0 + H_1(\gamma_0) \frac{\pi}{\sqrt{T}} \right) \xrightarrow{AS} \begin{pmatrix} \alpha_0 \\ 0 \\ 0 \end{pmatrix},$$

and

$$\begin{aligned} \sqrt{T} \begin{pmatrix} \hat{\alpha}(\gamma) - \alpha_0 \\ \hat{\theta}(\gamma) - \theta_0 \\ \hat{\varphi}(\gamma) \end{pmatrix} &= \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'e}{\sqrt{T}} + \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'H_1(\gamma_0)}{T} \pi - J_1 \pi \\ &\Rightarrow Q(\gamma, \gamma)^{-1} (\Psi(\gamma) + (Q(\gamma, \gamma_0) - Q(\gamma, \gamma)) J_1 \pi). \end{aligned}$$

When γ_0 is estimated, under H_0

$$\sqrt{T} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\theta} - \theta_0 \\ \hat{\varphi} \end{pmatrix} \xrightarrow{d} Q(\gamma^*, \gamma^*)^{-1} (\Psi(\gamma^*) + (Q(\gamma^*, \gamma_0) - Q(\gamma^*, \gamma^*)) J_1 \pi).$$

Thus, the asymptotic distributions of $\hat{\alpha}$, $\hat{\theta}$ and $\hat{\varphi}$ are non-normal under H_0 .

3 Test statistics and their asymptotic behavior

We start by deriving asymptotic distributions of the classical test statistics under a drifting parameter DGP. By ‘‘classical’’ we mean the standard Wald, Likelihood Ratio and Lagrange Multiplier tests constructed for fixed values of γ and γ_C when the model is linear with respect to the other parameters, with these values set to the estimates $\hat{\gamma}$ and $\hat{\gamma}_C$. This approach is motivated by the non-differentiability of $h_t(\gamma)$ in our leading SETAR example in which the standard test statistics so constructed are asymptotically χ^2 when γ_0 is well identified even though it is estimated.

When γ is fixed, the Wald statistic equals

$$\begin{aligned} W(\gamma) &= \frac{1}{ASR(\gamma)} \left(\left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'e}{\sqrt{T}} + \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'H(\gamma_0)}{T} J_1 \pi \right)' \\ &\quad \times J_2 \left(J_2' \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} J_2 \right)^{-1} J_2 \\ &\quad \times \left(\left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'e}{\sqrt{T}} + \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \frac{H(\gamma)'H(\gamma_0)}{T} J_1 \pi \right). \end{aligned}$$

When γ_0 is estimated, the Wald statistic W is asymptotically distributed as follows:

$$W \xrightarrow{d} \frac{1}{\sigma^2} \zeta'_W \zeta_W,$$

where

$$\zeta_W = (J_2' Q(\gamma^*, \gamma^*)^{-1} J_2)^{-\frac{1}{2}} J_2' Q(\gamma^*, \gamma^*)^{-1} (\Psi(\gamma^*) + Q(\gamma^*, \gamma_0) J_1 \pi).$$

When γ and γ_C are fixed, the Likelihood Ratio statistic equals

$$\begin{aligned} LR(\gamma, \gamma_C) &= T \log \left(\left(\frac{H(\gamma)'e}{\sqrt{T}} + \frac{H(\gamma)'H(\gamma_0)}{T} J_1 \pi \right)' \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \right. \\ &\quad \times \left. \left(\frac{H(\gamma)'e}{\sqrt{T}} + \frac{H(\gamma)'H(\gamma_0)}{T} J_1 \pi \right) \right)^{-1} \left(\left(\frac{H(\gamma_C)'e}{\sqrt{T}} + \frac{H(\gamma_C)'H(\gamma_0)}{T} J_1 \pi \right)' \right. \\ &\quad \times \left. J_{x1} \left(J'_{x1} \frac{H(\gamma_C)'H(\gamma_C)}{T} J_{x1} \right)^{-1} J'_{x1} \left(\frac{H(\gamma_C)'e}{\sqrt{T}} + \frac{H(\gamma_C)'H(\gamma_0)}{T} J_1 \pi \right) \right). \end{aligned}$$

When γ_0 is estimated, the Likelihood Ratio statistic LR has the following asymptotic distribution, also shared by k_2 times the classical F statistic:

$$LR \xrightarrow{d} \frac{1}{\sigma^2} (\zeta'_{CLR} \zeta_{CLR} - \zeta'_{LR} \zeta_{LR}),$$

where

$$\begin{aligned} \zeta_{LR} &= Q(\gamma^*, \gamma^*)^{-\frac{1}{2}} (\Psi(\gamma^*) + Q(\gamma^*, \gamma_0) J_1 \pi), \\ \zeta_{CLR} &= (J'_{x1} Q(\gamma_C^*, \gamma_C^*) J_{x1})^{-\frac{1}{2}} J'_{x1} (\Psi(\gamma_C^*) + Q(\gamma_C^*, \gamma_0) J_1 \pi). \end{aligned}$$

When γ and γ_C are fixed, the Lagrange Multiplier statistic equals

$$\begin{aligned} LM(\gamma, \gamma_C) &= \frac{1}{ASR_C(\gamma_C)} \left(e + \frac{H(\gamma_0)}{\sqrt{T}} J_1 \pi \right)' \\ &\quad \times \left(I - \frac{H_{x1}(\gamma_C)}{\sqrt{T}} \left(\frac{H_{x1}(\gamma_C)'H_{x1}(\gamma_C)}{T} \right)^{-1} \frac{H_{x1}(\gamma_C)'}{\sqrt{T}} \right) \frac{H(\gamma)}{\sqrt{T}} \left(\frac{H(\gamma)'H(\gamma)}{T} \right)^{-1} \\ &\quad \times \frac{H(\gamma)'}{\sqrt{T}} \left(I - \frac{H_{x1}(\gamma_C)}{\sqrt{T}} \left(\frac{H_{x1}(\gamma_C)'H_{x1}(\gamma_C)}{T} \right)^{-1} \frac{H_{x1}(\gamma_C)'}{\sqrt{T}} \right) \left(e + \frac{H(\gamma_0)}{\sqrt{T}} J_1 \pi \right), \end{aligned}$$

and has the following asymptotic distribution when γ_0 is estimated:

$$LM \xrightarrow{d} \frac{1}{\sigma^2} \zeta'_{LM} \zeta_{LM},$$

where

$$\zeta_{LM} = Q(\gamma^*, \gamma^*)^{-\frac{1}{2}} \left(\begin{array}{l} \Psi(\gamma^*) + Q(\gamma^*, \gamma_0) J_1 \pi - Q(\gamma^*, \gamma_C^*) J_{x1} (J'_{x1} Q(\gamma_C^*, \gamma_C^*) J_{x1})^{-1} \\ \times J'_{x1} (\Psi(\gamma_C^*) + Q(\gamma_C^*, \gamma_0) J_1 \pi) \end{array} \right).$$

The null distributions for the three tests are nonstandard, and resemble asymptotic distributions under local alternatives. Moreover, they depend on a host of nuisance parameters and consistently non-estimable value of π , thus tabulation of critical values is problematic. The distributions are in general different from each other. It can be shown though that $W \geq LR$, which implies that the Wald test rejects more often than the Likelihood Ratio test.

It is also interesting to investigate the tests that are designed for the case of complete non-identification under H_0 . Such tests are developed in Davies (1987), Andrews and Ploberger (1994) and Hansen (1996), among others. We will consider the most popular class of tests, the ones with the supremum functional; the others can be treated similarly. In the present context these tests may be applied in two variations.

In the first variation, the researcher neglects some identifiability of γ_0 and treats it as completely non-identified, effectively testing the hypothesis $H'_0 : \theta_0 = \varphi_0 = 0$. In this case, the sup-Wald statistic $\sup W_1$, sup-Likelihood Ratio statistic $\sup LR_1$ and sup-Lagrange Multiplier statistic $\sup LM_1$ have the following asymptotic behavior under H_0 :

$$T_1 \xrightarrow{d} \frac{1}{\sigma^2} \sup_{\gamma \in \Gamma} \zeta_1(\gamma)' \zeta_1(\gamma),$$

where T_1 is $\sup W_1$, $\sup LR_1$, or $\sup LM_1$, and

$$\zeta_1(\gamma) = (J'_{12}Q(\gamma, \gamma)^{-1}J_{12})^{-\frac{1}{2}} J'_{12}Q(\gamma, \gamma)^{-1} (\Psi(\gamma) + Q(\gamma, \gamma_0)J_1\pi)$$

(cf. Hansen (1996, Theorem 1)).

In the second variation, the researcher acknowledges the smallness of θ_0 and simply sets it equal to zero. In this case, the sup-Wald, sup-Likelihood Ratio and sup-Lagrange Multiplier statistics $\sup W_2$, $\sup LR_2$ and $\sup LM_2$ have the following asymptotic behavior under H_0 :

$$T_2 \xrightarrow{d} \frac{1}{\sigma^2} \sup_{\gamma \in \Gamma} \zeta_2(\gamma)' \zeta_2(\gamma),$$

where T_2 is $\sup W_2$, $\sup LR_2$, or $\sup LM_2$, and

$$\zeta_2(\gamma) = \left(J'_2 J_{x2} (J'_{x2} Q(\gamma, \gamma) J_{x2})^{-1} J'_{x2} J_2 \right)^{-\frac{1}{2}} J'_2 J_{x2} (J'_{x2} Q(\gamma, \gamma) J_{x2})^{-1} J'_{x2} (\Psi(\gamma) + Q(\gamma, \gamma_0) J_1 \pi).$$

4 Threshold autoregression

To assess the size distortions provided by the tests under study in the situation with weak identifiability, we consider our leading example of a two regime self-exciting threshold first-order autoregression, reparameterized for the sake of symmetry:

$$y_t = \varepsilon_t + \begin{cases} (\mu_0 - \Delta\mu/2) + (\rho_0 - \Delta\rho/2) y_{t-1} & \text{if } y_{t-1} \leq \gamma_0, \\ (\mu_0 + \Delta\mu/2) + (\rho_0 + \Delta\rho/2) y_{t-1} & \text{if } y_{t-1} > \gamma_0. \end{cases}$$

The null hypothesis is $H_0 : \Delta\rho = 0$. When $\Delta\mu$ is big, the conventional asymptotic theory is expected to yield adequate approximation to distributions of standard test statistics; when $\Delta\mu = 0$, it is appropriate to apply the Davies–Andrews&Ploberger–Hansen statistics (in which case it is a test for linearity), and when $\Delta\mu$ is small but non-zero, the actual

distributions of test statistics of both types may deviate significantly from predictions of the standard asymptotic theory.

We set $T = 300$, $\mu_0 = \rho_0 = 0$, $\gamma_0 = 0$, and $\Upsilon = [0.1, 0.9]$, where Υ is the image of the transformation by the EDF of y_t with domain Γ (for more details about Υ , see Hansen (1996, p. 420)). The number of repetitions is 10,000 implying that the standard error is 0.22% for the actual size of 5%, and does not exceed 0.46% for a maximal reported actual size. We use the Hansen (1997) procedure to compute asymptotic p-values for sup tests, while the critical value for classical tests is the 95% quantile of chi-squared distribution. All computations are performed using GAUSS, version 4.0.23.

Figure 1 presents distributions of threshold estimates in the unrestricted and restricted models when $\Delta\mu = 0.2$. These distributions are quite dispersed, and look like a mixture of two distributions, one of which is tightly concentrated around γ_0 , and the other is completely uninformative about γ_0 . The distribution of the unrestricted threshold estimates is even triple-peaked, although it is only for a small range of values of $\Delta\mu$. The correlation coefficient between the two threshold estimates equals 0.55 and is nearly invariant to the value of $\Delta\mu$.

Figure 2 depicts the actual sizes for the nominal size of 5% when $\Delta\mu$ varies between 0 and 1, with a fine step of 0.02 when $\Delta\mu$ is small, a medium step of 0.05 when $\Delta\mu$ is moderate, and a rough step of 0.25 when $\Delta\mu$ is large. The classical tests, except the LR , exhibit severe overrejection not only when $\Delta\mu$ is very small but also for its moderate values. In contrast, the classical LR statistic displays surprising robustness to the degree of identification, the actual size not exceeding 9% even under near non-identification. The T_1 tests that ignore some identifiability of γ_0 (we show the plots for only LR associated sup tests because the others are extremely close) work well only in a close neighborhood of non-identifiability, the size distortion beginning to rise sharply when $\Delta\mu$ exceeds 0.1 (so that π exceeds ≈ 2). The T_2 tests that remove the problematic term reject much less frequently than they should for a wide range of identification degrees, but exhibit serious overrejection when γ_0 is very well identified.

We also run several experiments to assess the robustness of sizes to parameter values. Starting from the DGP with basic parameter values, we change the threshold value γ_0 to 0.5, 1.0, 1.5; or the set Υ to $[0.05, 0.95]$, $[0.15, 0.85]$, $[0.2, 0.8]$; or the sample size T to 100, 200, 400, 500 simultaneously changing $\Delta\mu$ so that $\pi = \Delta\mu\sqrt{T}$ is held constant; or the persistence parameter ρ_0 to 0.3, 0.6, 0.9, 0.99. Most test statistics turn to be pretty robust for changes in all parameters, with the following notable exceptions. As Υ is shrinking, the sizes of the classical W and LM tests quickly drop and reach about 15% when $\Upsilon = [0.2, 0.8]$, while that of the classical LR test drops from 8% to 6%. Further, as the persistence parameter ρ_0 goes up above 0.6, the former two statistics exhibit even higher rejection rates which exceed 25%

when ρ_0 is 0.9 and 35% when ρ_0 is 0.99. The rejection rate of the LR test also increases, but only to 11% when ρ_0 is 0.9 and to 16%, when ρ_0 is 0.99. Similar tendencies, although to a much lesser degree, take place when the threshold value γ_0 is shifted.

Finally, we examine the power of the LR test to make sure that its favorable size properties do not come at the expense of power properties. We compute the 5% critical value from the simulated distribution of the LR statistic under the null with the basic parameter combination. We set the slope shift $\Delta\rho$ at 0.2, 0.4 and 0.8, and obtain pretty decent power values of 11.8%, 20.4% and 56.1%, respectively.

5 Concluding remarks

The standard asymptotic theory breaks down when a nuisance parameter is weakly identified under the null hypothesis. The classical asymptotic tests display significant size distortions, although simulations with a simple SETAR model show that the Likelihood Ratio test seems to be affected by the phenomenon in the least degree. The tests designed for the case of complete non-identification under the null work well only for a narrow range of parameter values. Constructing an asymptotically pivotal test statistic robust to the degree of identification remains a challenging task.

References

- Andrews, D.W.K. and W. Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383–1414.
- Davies, R.B., 1987, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 74, 33–43.
- Hansen, B., 1996, Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 413–430.
- Hansen, B., 1997, Approximate asymptotic p-values for structural change tests, *Journal of Business and Economic Statistics* 15, 60–67.
- 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 identification, *Econometrica* 68, 1055–1096.
- Tripathi, G., 1999, A matrix extension of the Cauchy–Schwarz inequality, *Economics Letters* 63, 1–3.

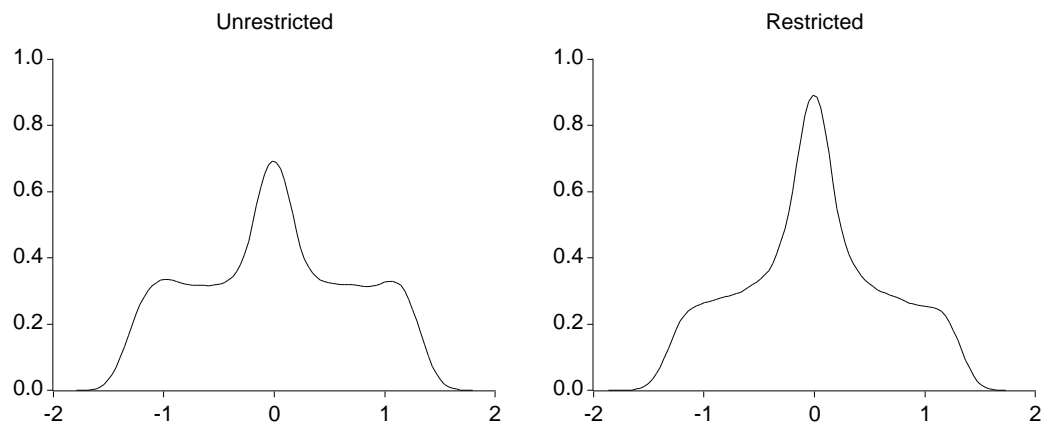


Figure 1. Distributions of threshold estimates in unrestricted and restricted SETAR models.

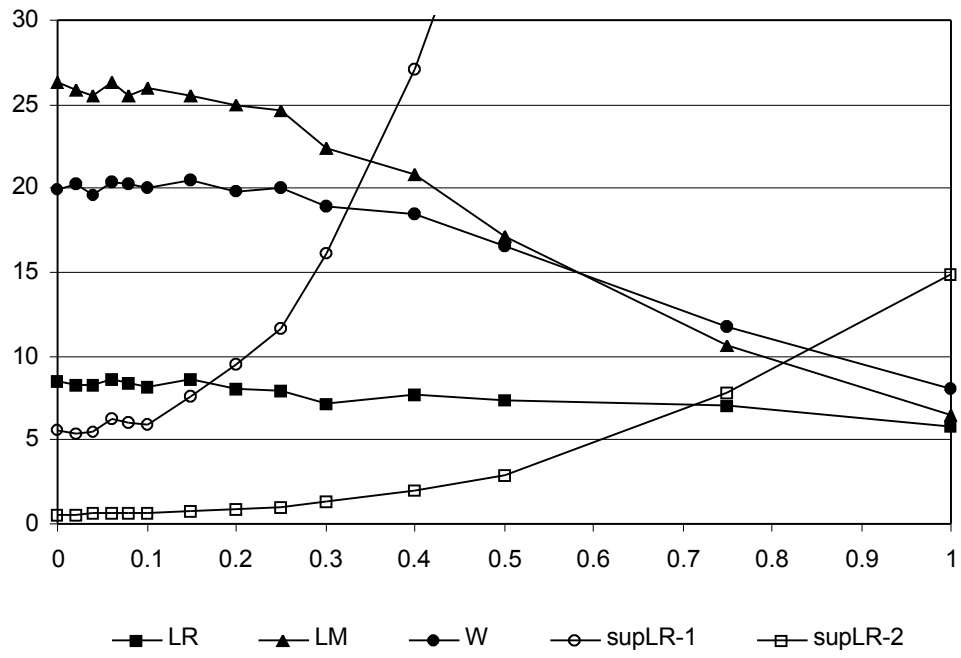


Figure 2. Actual test sizes as functions of mean shift in SETAR model.