Testing linearity against smooth transition autoregression using a parametric bootstrap

Joakim Skalin
Department of Economic Statistics
Stockholm School of Economics
P.O. Box 6501, SE-113 83 Stockholm, Sweden
E-mail: Joakim.Skalin@hhs.se

Stockholm School of Economics
Working Paper Series in Economics and Finance No. 276

Revised version

13 December 1998

Abstract

When testing the null hypothesis of linearity of a univariate time series against smooth transition autoregression (STAR), standard asymptotic distribution results do not apply since nuisance parameters in the model are unidentified under the null hypothesis. The prevailing test of Luukkonen, Saikkonen and Teräsvirta (1988) is based on a linearization, which may adversely affect its power. This paper discusses an alternative procedure, based on a parametric bootstrap of a likelihood ratio test statistic, and investigates its size and power properties by a small simulation study. The results, however, indicate that the power of the bootstrap test is inferior to that of the existing test.

Keywords. Linearity testing; smooth transition autoregression model; nuisance parameter; nonstandard testing problem; bootstrap test.

JEL Classification Codes: C12; C15; C22

Acknowledgments. The author wishes to thank Timo Teräsvirta for valuable discussions, suggestions and comments and Tor Jacobson for useful comments on an earlier version of the paper. The responsibility for all errors and shortcomings remains with the author. Financial support from the The Bank of Sweden Tercentenary Foundation is gratefully acknowledged.

1 Introduction

Testing hypotheses about parameters in econometric models when one or more nuisance parameters are not identified under the null hypothesis requires non-standard methods, since the standard distribution results for the classical test statistics (the likelihood ratio, Lagrange multiplier, and Wald tests) do not apply. The problem arises in many situations in applied work. For instance, when the null of linearity of a time series is tested against a smooth transition autoregressive (STAR) model (or against any one of a number of other nonlinear alternative specifications) some of the parameters are unidentified under the null hypothesis.

Among the different solutions to this general statistical problem suggested in the literature, the method proposed by Teräsvirta (1994) is the most commonly utilised one in STAR modelling. However, since this test procedure circumvents the identification problem by an appropriate linearisation (see, e.g., Luukkonen, Saikkonen and Teräsvirta (1988) for details) information about the nonlinear structure under the alternative is lost and the power may be adversely affected. If so, it might be worthwhile to consider other tests. While this test is based on standard asymptotic distribution theory, solutions suggested by other authors in practice rely on computational statistical methods for establishing critical values for the test statistics. In this spirit, Hansen (1996) developed a technique for establishing critical values for test statistics derived by Andrews and Ploberger (1994) but his method is computationally very demanding.

Desirable properties of any competing method include reasonable computational costs and, above all, at least as good power properties as the linearisation-based procedure mentioned above. The present paper is a first attempt at investigating whether bootstrapping a likelihood-ratio test of linearity against STAR meets these requirements and therefore merits further research.

The rest of the paper is organised as follows: In Section 2 the STAR model and the linearity test are introduced, the identification problem is discussed, and some previously suggested general solutions are briefly reviewed. Section 3 describes the bootstrap test and Section 4 the simulation study. Section 5 concludes.

2 Testing linearity against STAR

2.1 The STAR model

A smooth transition autoregression, STAR, model of order p is defined as

$$y_t = \theta' w_t + (\pi' w_t) F(y_{t-d}; \gamma, c) + u_t, \tag{1}$$

where $u_t \sim nid\left(0, \sigma_u^2\right)$, $\theta = (\theta_0, \theta_1, \ldots, \theta_p)'$, $\pi = (\pi_0, \pi_1, \ldots, \pi_p)'$, and $w_t = (1, y_{t-1}, \ldots, y_{t-p})'$. Model (1) is called a logistic smooth transition autoregression, LSTAR, model, if the transition function $F(y_{t-d}; \gamma, c)$ is defined as

$$F(y_{t-d}; \gamma, c) = \left(1 + \exp\left\{-\gamma\left(y_{t-d} - c\right)\right\}\right)^{-1}, \gamma > 0 \tag{2}$$

and an exponential, ESTAR, model, if

$$F(y_{t-d}; \gamma, c) = 1 - \exp\left\{-\gamma\left(y_{t-d} - c\right)^2\right\}, \gamma > 0. \tag{3}$$

Early discussions of smooth transition (auto)regression models can be found in Bacon and Watts (1971), Goldfeld and Quandt (1972) and Maddala (1977, p. 396). More recently, STAR models have received attention in both theoretical and applied work. Teräsvirta (1994) and Eitrheim and Teräsvirta (1996) present a modelling cycle for STAR models inluding testing linearity against STAR, specifying and estimating a STAR model if linearity is rejected, and evaluating the estimated model by a number of diagnostic tests. An overview of the modelling cycle, as well as applications and references, can be found in Teräsvirta (1998).

The first step of the modelling cycle, testing the null hypothesis of linearity against the proposed nonlinear model, is crucial; clearly, nonlinear modelling should not be attempted if a linear model is an adequate representation of the data. However, devising a test of linearity against the STAR model is complicated by the fact that (1) is not identified under the null hypothesis. The next subsection describes the problem of hypothesis testing when nuisance parameters are not identified under the null hypothesis.

2.2 The identification problem

Testing linearity against STAR in (1) with (2) or (3) amounts to testing $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$. However, under the null hypothesis, both π and c are nuisance parameters that can take any value without affecting the likelihood. Thus the model is identified only under the alternative but not under the null hypothesis and the standard χ^2 asymptotic theory of the three classical tests does not hold. The presence of an identification problem is also seen by noticing the fact that the testing problem could equally well be formulated in terms of the lag parameters of the 'nonlinear part' of the model, i.e., testing $H_0 : \pi = 0$ against $H_1 : \pi \neq 0$.

The general problem of hypothesis testing when a nuisance parameter vector, β, is not identified under the null hypothesis was first discussed by Davies (1977, 1987), who suggested that a test statistic, $S(\beta)$, be first derived under the assumption that the nuisance parameter vector is fixed, $\beta = \beta^*$. The actual test statistic is then defined as $\sup_{\beta^* \in \mathcal{B}} S(\beta^*)$ (assuming a right-tailed test). Covering the 'worst case', this procedure is clearly conservative. The first econometric application of this idea was Watson and Engle (1981). The asymptotic null distribution of the supremum test statistic is not generally known analytically.

Luukkonen, Saikkonen and Teräsvirta (1988) and Saikkonen and Luukkonen (1988) suggested a technique that may be viewed as being similar in spirit to the solution of Davies (1977). Here, the transition function is replaced by a Taylor approximation around H_0 and the model is reparameterised in order to obtain an auxiliary linear regression where certain parameters equal zero if $\gamma = 0$, thereby circumventing the identification problem. This test has been used in a number of applied studies (see Teräsvirta (1998) for examples). In a comparative simulation study in Hansen (1996) the Taylor approximation based test performed well against a self-exciting threshold autoregression (SETAR) model.

Andrews and Ploberger (1994) derived optimal versions of the three classical tests for a situation where one of the nuisance parameters affects the likelihood only under the alternative hypothesis. However, these average exponential tests

require that the investigator specifies a weight function over the possible values of the unidentified nuisance parameter. From a Bayesian point of view the weight function may be seen as a prior, and the tests are asymptotically equivalent to Bayesian posterior odds ratios. Andrews and Ploberger do not discuss how critical values for the test statistics should be obtained in practice.

Hansen (1996) developed a procedure for computing simulated critical values for, e.g., the Davies and the Andrews and Ploberger test statistics. If the nuisance parameter that is not identified under the null belongs to a continuous parameter space, the investigator in practice has to select a number of discrete values for the parameter. In every replicate of the simulation a test statistic is then calculated given each one of these parameter values, and one of the methods suggested by Davies or Andrews and Ploberger is applied to arrive at a single test statistic for the replicate. The decision is then based upon the comparison of the original test statistic, in the form of a p-value, to the simulated distribution. The combination of simulating the distribution and the need for selecting a set of values for the nuisance parameter makes the method computationally intensive. Furthermore, Hansen (1996) explicitly discussed only the case where *one* nuisance parameter is unidentified under the null hypothesis. In the STAR case, whichever way the test is parameterised, the testing problem generally involves at least two such parameters. In practice, a generalisation to more than a single nuisance parameter would considerably increase the already substantial computational burden. It should also be noted that the power of the test is dependent on the selected set of values for the nuisance parameter (or parameters).

Given the statistical problems briefly described above and the increasing interest in simulation-based methods, reflected in part by the works cited above, it seems natural to consider using a bootstrap procedure to establish the empirical distribution of a linearity test. In the next section a bootstrap test that should be less computationally intensive than the one described in the previous paragraph is outlined, and its size and power properties are discussed in the subsequent section.

3 The bootstrap test

It is possible to construct a parametric bootstrap likelihood ratio test of linearity against STAR. The empirical distribution for the test statistic is established through resampling from time series generated using the estimates of the parameters under the null hypothesis and normally distributed random errors whose variance equals the estimated residual variance of the model. For a general discussion of bootstrap tests with examples and applications the reader is referred to, e.g., Davidson and MacKinnon (1996).

In order to obtain an empirical distribution of the test statistic, model (1) is assumed to be completely specified under the null and under the alternative, i.e., the parameters p (the autoregressive order), the form of the transition function F (LSTAR (2) or ESTAR (3)), and d (the delay parameter of the transition function) are assumed to be known and only the values of the parameters in the vector $\phi = \left[\theta', \pi', \gamma, c\right]'$, and σ_u^2, are unknown. In practice, neither the form of the transition function nor the value of the delay parameter d is typically known and has to be selected using the data. However, before considering the

test in this more complicated situation, its performance in the simplest case has to be investigated. The sample data set is denoted $y = [y_1, \ldots, y_T]'$. The test procedure consists of the following steps:

1. Estimate the model under H_0, i.e. a linear AR model. The estimates are denoted $\widetilde{\theta}$ and $\widetilde{\sigma}_u^2$ and the value of the likelihood function is L_{max}^0.
2. Estimate the model under H_1, i.e. the STAR model. The estimates are denoted $\hat{\phi} = \left[\hat{\theta}', \hat{\pi}', \hat{\gamma}, \hat{c}\right]'$ and $\hat{\sigma}_u^2$. The value of the likelihood function is L_{max}^1.
3. Compute the value of the LR statistic $\widehat{\tau} = -2\log\left(L_{max}^0/L_{max}^1\right)$.
4. Generate $T_B R$ pseudo random numbers, $u_{r,t}^* \sim N\left(0, \tilde{\sigma}_u^2\right), t = 1, \ldots, T_B, r = 1, \ldots R, T_B < T$.
5. Generate R time series of length T_B using the estimated model under H_0, $y_{r,t}^* = \tilde{\theta}' w_t^* + u_{r,t}^*$.
6. For each time series, $r = 1, \ldots R$, estimate the model under H_0, i.e. a linear AR model. The estimates are denoted $\tilde{\theta}^*$ and $\widetilde{\sigma}_{r,u}^{*2}$, and the value of the likelihood function is $L_{r,max}^{*0}$.
7. For each time series, $r = 1, \ldots R$, estimate the model under H_1, i.e. the STAR model. The estimates are denoted $\hat{\phi}_r^* = \left[\hat{\theta}_r^{*\prime}, \hat{\pi}_r^{*\prime}, \hat{\gamma}_r^*, \hat{c}_r^*\right]'$ and $\hat{\sigma}_{r,u}^{*2}$. The value of the likelihood function is $L_{r,max}^{*1}$.
8. For each time series, $r = 1, \ldots R$, compute the value of the LR statistic $\tau_r^* = -2\log\left(L_{r,max}^{*0}/L_{r,max}^{*1}\right)$.
9. Compute the estimated bootstrap p-value function (see, e.g., Davidson and MacKinnon, 1996) as the ratio

$$p^*\left(\hat{\tau}\right) = \frac{\#\left(\tau_r^* > \hat{\tau}\right)}{R}$$

   where $\#\left(\tau_r^* > \hat{\tau}\right)$ is the number of times over the R series that $\tau_r^* > \hat{\tau}$, $r = 1, \ldots, R$.
10. If $p^*\left(\hat{\tau}\right) < \alpha$, where α is the selected significance level, the null hypothesis of linearity is rejected.

Note that $T_B < T$, i.e. the bootstrap replicate series is shorter than the original time series. This choice is based on the theoretical considerations in Bickel, Götze and van Zwet (1997); asymptotically, $T_B/T \to 0$ as $T_B, T \to \infty$. (See also Rydén, Teräsvirta, and Åsbrink (1998) for a discussion and an application.)

In practice it may happen that the nonlinear least squares estimation of the STAR model in step 2 or 7 above converges to a local optimum rather than the global, or simply fails to converge at all. Convergence to a local optimum due to a flat likelihood function cannot be controlled, but since only the estimated value of the likelihood function, and not the estimated parameter

vector, enters the test statistic, the effects of a flat likelihood should not be too severe. If the numerical optimisation frequently fails to converge at all, however, the results become difficult to interpret. To mitigate this problem, the optimisation is restarted with a new set of initial values for the parameters whenever the first attempt fails to converge within a certain limit on computing time or on the number of iterations. Only if optimisation starting from, say, three or four different initial parameter vectors fails to yield convergence the simulation replicate (if the problem appears at step 2) or the bootstrap replicate (if the problem occurs at step 7) is dropped. In the simulations reported below, the problems of non-convergence turn out to be very small.

4 Size and power properties

In this section, the power and size properties of the proposed procedure are illustrated by means of a small Monte Carlo study. Since the nonlinear least squares estimation of the STAR model in the second and seventh steps of the test procedure described in the previous section is rather time-consuming, and given scarce computing resources, only a very limited simulation study has been possible at this stage, and the results should be regarded as indicative.

4.1 The simulation setup

In the size simulations, the data are generated by

$$y_t = 1.3y_{t-1} - 0.5y_{t-2} + u_t \tag{4}$$

where u_t is a pseudo random number, $u_t \sim nid\left(0, \sigma_u^2\right)$, with $\sigma_u^2 = 0.0004$ or $\sigma_u^2 = 0.001$. In the power simulations, data are generated either by the LSTAR model

$$\begin{aligned} y_t &= 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2}) \\ &\quad \times \left(1 + \exp\left\{-\gamma\left(y_{t-1} - 0.02\right)\right\}\right)^{-1} + u_t \end{aligned} \tag{5}$$

where $\gamma = 20$ or $\gamma = 100$ (cf Teräsvirta 1994, eq. 4.1-4.2) or by the ESTAR model

$$\begin{aligned} y_t &= 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2}) \\ &\quad \times \left(1 - \exp\left\{-\gamma\left(y_{t-1} - 0.02\right)^2\right\}\right) + u_t \end{aligned} \tag{6}$$

with $\gamma = 100$ or $\gamma = 1000$ (cf Teräsvirta 1994, eq. 4.1 and 4.6). In both cases, $u_t \sim nid\left(0, \sigma_u^2\right)$, with $\sigma_u^2 = 0.0004$. Two different sample sizes are used, $T = 100$ (with $T_B = 90$) and $T = 500$ (with $T_B = 450$). As noted above, due to computation time constraints, the simulation study is rather limited in scope: $R = 200$ bootstrap replicates are used throughout, and the simulations are all based on 500 replicates. For comparison, the empirical size and power of the linearity test of Teräsvirta (1994) have been calculated for the same data generating processes. In these simulations, however, 10000 replicates were used. A nominal significance level of $\alpha = 0.05$ is applied throughout.

The simulation study is programmed in *Gauss for Windows* (using the *Optmum* package for all numerical optimisation) and executed on personal computers. Since the equipment used varies considerably with respect to, e.g., processor

type and speed, no systematic evaluation of the computing time is made, but run times for one simulation experiment range approximately from 30 hours to 200 hours (on PCs with frequencies between 200 and 350 MHz).

4.2 Simulation results

The results of the simulations are given in Tables 1-4. Tables 1 and 2 indicate that the bootstrap test is fairly well-sized, making a power simulation meaningful. The Taylor approximation based test is slightly conservative. The empirical power of the bootstrap test is lower than or equal to that of the Taylor approximation based test in all investigated cases except for the ESTAR model with the sharper transition where the bootstrap test has better power. On the other hand, the ESTAR model with the smoother transition is the case that favours the auxiliary regression based test the most. As $\gamma \to \infty$ the ESTAR model approaches a linear model (with probability one) and the ESTAR realisations become more difficult to distinguish from linear ones since the 'nonlinearity' involves a very small number of observations in the sample. This is illustrated by the power results for the ESTAR data generating process with $\gamma = 1000$; with a very sharp transition, there is power to be gained by not omitting information about the nonlinear structure through linearisation. The problems of non-convergence seem to be manageable in the present setting.

5 Conclusions

The simulation study of this paper suggests that the bootstrap likelihood ratio test of linearity against STAR has good size properties but is generally less powerful than the auxiliary regression based test suggested by Luukkonen, Saikkonen and Teräsvirta (1988). The simulation results concerning the latter test accord well with previous studies. Thus the simulations do not indicate that the latter should be abandoned in favour of the bootstrap test. Furthermore, it should be noted that only testing against a fully known alternative model has been considered here. The auxiliary regression based test has the advantage that it is generally powerful against LSTAR and ESTAR simultaneously if the order of the Taylor expansion is at least two. This property is a useful one in STAR model building; see, for example, Teräsvirta (1994).

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.

Bacon, D. W., and D. G. Watts (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58, 525-534.

Bickel, P. J., F. Götze and W. R. van Zwet (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7, 1-31.

Davidson, R. and J. G. MacKinnon (1996). The size distortion of bootstrap tests. Queen's University, Institute for Economic Research Discussion Paper No. 936.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64, 247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74, 33-44.

Eitrheim, Ø. and T. Teräsvirta (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74, 59-75.

Goldfeld, S. M. and R. Quandt (1972). *Nonlinear methods in econometrics*. Amsterdam: North-Holland.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64, 413-430.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75, 491-499.

Maddala, D. S. (1977). *Econometrics*. New York: McGraw-Hill.

Rydén, T., T. Teräsvirta, and S. Åsbrink (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13, 217-244.

Saikkonen, P. and R. Luukkonen (1988). Lagrange multiplier tests for testing non-linearities in time series models. *Scandinavian Journal of Statistics*, 15, 55-68.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89, 208-218.

Teräsvirta, T. (1998). Modeling economic relationships with smooth transition regressions. In Ullah, A. and D. E. A. Giles (eds.), *Handbook of Applied Economic Statistics*, 507-552. New York: Dekker.

Watson, M. W. and R. F. Engle (1985). Testing for regression coefficient stability with a stationary AR(1) alternative. *Review of Economics and Statistics*, 67, 341-346.

Table 1. Empirical size for sample size = 100, and numbers of simulation replicates discarded due to non-convergent estimation (in square brackets)

Data generating process		Bootstrap test		Approximation based test
		Alternative model LSTAR (eq. 5)	Alternative model ESTAR (eq. 6)	
AR (eq. 4)	$\sigma_u^2 = 0.0004$	0.058 [0]	0.038 [0]	0.039
	$\sigma_u^2 = 0.01$	0.052 [6]	0.052 [6]	0.039

Table 2. Empirical size for sample size = 500, and numbers of simulation replicates discarded due to non-convergent estimation (in square brackets)

Data generating process		Bootstrap test		Approximation based test
		Alternative model LSTAR (eq. 5)	Alternative model ESTAR (eq. 6)	
AR (eq. 4)	$\sigma_u^2 = 0.0004$	0.050 [0]	0.056 [0]	0.041
	$\sigma_u^2 = 0.01$	0.048 [2]	0.054 [11]	0.041

Table 3. Empirical power for sample size = 100

Data generating process		Alternative model			
		LSTAR (eq. 5)		ESTAR (eq. 6)	
		Bootstrap test	Approx. based test	Bootstrap test	Approx. based test
LSTAR (eq. 5)	$\gamma = 20$	0.328	0.391		
	$\gamma = 100$	0.848	0.947		
ESTAR (eq. 6)	$\gamma = 100$			0.754	0.907
	$\gamma = 1000$			0.942	0.873

Table 4. Empirical power for sample size = 500

Data generating process		Alternative model			
		LSTAR (eq. 5)		ESTAR (eq. 6)	
		Bootstrap test	Approx. based test	Bootstrap test	Approx. based test
LSTAR (eq. 5)	$\gamma = 20$	0.802	0.994		
	$\gamma = 100$	0.978	1.000		
ESTAR (eq. 6)	$\gamma = 100$			0.996	1.000
	$\gamma = 1000$			1.000	1.000