Nonlinear dynamics and smooth transition models

Andrés González Gómez

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen vid Handelshögskolan i Stockholm framläggs för offentlig granskning Onsdagen den 22 september 2004, kl. 12.30 i sal 750, Handelshögskolan, Sveavägen 65.

Nonlinear dynamics and smooth transition models

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

EFI Mission
EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics.
EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Organization
The research activities are organized in twenty Research Centers within eight Research Areas. Center Directors are professors at the Stockholm School of Economics.

ORGANIZATION AND MANAGEMENT	
Management and Organisation; (A)	Prof Sven-Erik Sjöstrand
Center for Ethics and Economics; (CEE)	Adj Prof Hans de Geer
Center for Entrepreneurship and Business Creation; (E)	Prof Carin Holmquist
Public Management; (F)	Prof Nils Brunsson
Information Management; (I)	Prof Mats Lundeberg
Center for People and Organization; (PMO)	Prof Jan Löwstedt
Center for Innovation and Operations Management; (T)	Prof Christer Karlsson
ECONOMIC PSYCHOLOGY	
Center for Risk Research; (CFR)	Prof Lennart Sjöberg
Economic Psychology; (P)	Prof Guje Sevón
MARKETING	
Center for Consumer Marketing; (CCM)	Acting Prof Magnus Söderlund
Center for Information and Communication Research; (CIC)	Adj Prof Bertil Thorngren
Marketing, Distribution and Industrial Dynamics; (D)	Prof Björn Axelsson
ACCOUNTING, CONTROL AND CORPORATE FINANCE	
Accounting and Managerial Finance; (B)	Prof Lars Östman
Managerial Economics; (C)	Prof Peter Jennergren
FINANCE	
Finance; (FI)	Prof Clas Bergström
ECONOMICS	
Center for Health Economics; (CHE)	Prof Bengt Jönsson
International Economics and Geography; (IEG)	Prof Mats Lundahl
Economics; (S)	Prof Lars Bergman
ECONOMIC STATISTICS	
Economic Statistics; (ES)	Prof Anders Westlund
LAW	
Law; (RV)	Prof Erik Nerep

Chairman of the Board: Prof Carin Holmquist. *Director:* Associate Prof Filip Wijkström.

Adress
EFI, Box 6501, S-113 83 Stockholm, Sweden • Internet: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

Nonlinear dynamics and smooth transition models

Andrés González Gómez

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

Dissertation for the Degree of Doctor of Philosophy, Ph.D
Stockholm School of Economics 2004

ISBN NR 91-7258-645-1

Keywords:
Smooth Transition models, Hypothesis Testing

Distributed by:
EFI, Stockholm School of Economics
Box 6501, S-113 83 Stockholm, Sweden.
`www.hhs.se/efi`

Printed by:
Elanders Gotab, Stockholm 2004

A mis papás
por su ejemplo

Contents

Preface

During this last four years, I have come across many people who have increased my understanding and knowledge of econometrics. I am particularly grateful to my advisor Timo Teräsvirta whose excellent guidance has smoothed out my transition to econometric research. Our weekly meetings have been a fruitful experience in which he has given me invaluable insights into the world of econometrics.

I wish to express my sincere gratitude to my co-authors Dick van Dijk and Changli He whose contributions have greatly improved the contents of this thesis.

I would also like to thank the Stockholm School of Economics and the people at the Department of Economic Statistics for a pleasant environment and the necessary resources for developing my research. I must mention Sune Karlsson, not only for his advice on several econometric issues but also for creating STATBAT, the computer programme that has allowed me to run so many simulations so many times over. Special thanks also to my fellow PhD students for their useful comments during seminars and coffee breaks and to Pernilla Watson for patiently helping me with my practical day-to-day problems.

My PhD has also been enriched by the Stockholm University, Joensuu University and, finally, by Université de Toulouse I, where I was a visiting graduate student at the Department of Economics and whose kind hospitality is gratefully acknowledged.

I would like to thank my friends at the PhD program and the small Latin-American community for making my life in Sweden such an enjoyable experience. Most of all, I thank my wife Carolina, who has given me her loving company and unconditional support specially at times of falling enthusiasm.

Last but not least, I would like to thank the Central Bank of Colombia, Banco de la República, Jan Wallander's and Tom Hedelius's Foundation Grants No. J99/37 and J02-35, and the Economic Research Institute EFI at the Stockholm School of Economics for financial support.

Part I

Summary of Thesis

Introduction and summary

During the last few years nonlinear models have been a very active area of econometric research: new models have been introduced and existing ones generalized. To a large extent, these developments have concerned models in which the conditional moments are regime-dependent. In such models, the different regimes are usually linear and the change between them is governed by an observable or unobservable variable. For example, when only the conditional mean is regime-dependent the model can be written as follows:

$$y_t = (1 - g(q_t))x_t'\beta_1 + g(q_t)x_t'\beta_2 + e_t \tag{1}$$

where y_t is the dependent variable, x_t is a $(1 \times k)$ vector of explanatory variables, β_1 and β_2 are parameter vectors and e_t the errors.

The main feature in (1) is the transition function $g(q_t)$ which is a function of the transition variable q_t and normally takes values in the interval [0,1]. From (1) the role of the transition function is clear: the conditional mean of y_t equals $x_t'\beta_1$ when $g(q_t) = 0$ and $x_t'\beta_2$ when $g(q_t) = 1$. These specifications can be useful in situations in which it is suspected that the behaviour of the dependent variable may vary between regimes. A classical example can be found the business cycle literature where it is argued that contractions in the economy are not only more violent but also short-lived than expansions. Unemployment, which tends to rise faster during recessions than decline during booms, constitute another example.

Two of the most popular regime-dependent models are the smooth transition and the threshold model. In both models q_t is observable but the specification of $g(q_t)$ is different. Particularly, in the smooth transition model $g(q_t)$ is a continuous function of q_t whereas in the threshold model it is an indicator function that takes value one when q_t is above some level and zero otherwise. Conceptually, the difference between these models is that while in the threshold model each realization of y_t comes from a particular extreme regime, in the smooth transition model there are infinite number of regimes within the two extremes: $g(q_t) = 0$ and 1.

This difference between the smooth transition and the threshold model has also consequences for the statistical theory behind them. In particular, the discontinuity of $g(q_t)$ in the threshold model carries to the likelihood function, which implies that the standard statistical theory is not applicable and more elaborate methods are needed. Hansen (1996,1999) have developed appropriate statistical methods for this situation. In the smooth transition model, on the contrary, much of the statistical inference including parameter estimation can be carried out using standard statistical methods. Perhaps this technical advantage of the smooth transition model has allowed extensions of the model to many directions. A survey by van Dijk, Teräsvirta, and Franses (2002) presents a summary of many of the recent developments.

One of the factors that has influenced the development of nonlinear models are improvements in computer technology. They have not only permitted an introduction of more complex models but have also allowed the use of computer-intensive methods in hypothesis testing. This is particularly important in nonlinear models because there these methods have proved to be practical in testing statistical hypothesis such as linearity and parameter constancy. In general, these testing situation are not trivial and their solution often requires computer-intensive methods. In particular, bootstrapping and Monte Carlo testing are now commonly used; see Dufour and Khalaf (2001).

Bootstrapping and Monte Carlo testing are computationally intensive methods in which the empirical distribution function of a test statistic under the null hypothesis is constructed by simulation. It has been shown that in many cases small-sample properties of a test can be improved by using these techniques. The main difference between bootstrapping and Monte Carlo testing is that in the latter the small sample distribution of a given test statistic is obtained by simulation. That is, in Monte Carlo testing, both the sample size and the nuisance parameters under the null hypothesis are dealt with by the simulations. Monte Carlo testing also allows the introduction of new test statistics for which the small-sample distribution is unknown and probably analytically intractable.

In this thesis the smooth transition model is used in different ways. In the first chapter, a vector smooth transition model is used as a device for deriving a test for parameter constancy in stationary vector autoregressive models. In the second chapter we introduce a panel model whose parameters can change in a smooth fashion between regimes as a function of an exogenous variable. The method is used to investigate whether financial constraints affect firms' investment decisions. The third chapter is concern with linearity testing in smooth transition models. New tests are introduced and Monte Carlo testing

techniques are shown to be useful in achieving control over the size of the test. Finally, the last chapter is devoted to the Smooth Permanent Surge model. This is a nonlinear moving average model in which a shock can have transitory or permanent effects depending on its sign and magnitude. Test for linearity and random walk hypothesis are introduced. A more detail description of the chapters follows.

Chapter 1. Testing parameter constancy in stationary vector autoregressive models against continuous change[1]

In this chapter we use a vector smooth transition model in which the transition variable is time, as a device for deriving a test for parameter constancy in stationary vector autoregressive models. The model under the alternative hypothesis is a generalized version of (1) in which y_t is a vector of endogenous variables and x_t contains lags of y_t. The transition function is specified as in Jansen and Teräsvirta (1996) and the transition variable is always time, either $q_t = t$ or, standardized between zero and one as $q_t = t/T$ where T is the number of observations. The test of parameter constancy is derived using the approach proposed by Luukkonen, Saikkonen, and Teräsvirta (1988). This means that we generalize the single-equation parameter constancy test of Lin and Teräsvirta (1994) to the vector case. Such a generalization will be useful in building vector smooth transition autoregressive models, an area where further research is needed.

The asymptotic distribution of the test statistic is derived and its small-sample properties are investigated by simulation. However, as is often the case in vector autoregressive models, the asymptotic theory is not a reliable guide in small samples. This is particularly clear when the sample size compared to the number of parameters under the alternative is small. Monte Carlo experiments also show that by using the empirical distribution of the test, computed by bootstrapping, it is to a large extent possible to correct size distortion that may occur in small samples. The performance of the tests is compared to that of generalized Chow-tests and found satisfactory in terms of both size and power.

[1] This is a joint work with Timo Teräsvirta and Changli He.

Chapter 2. Panel smooth transition regression model and an application to investment under credit constraints[2]

This chapter describes another generalization of the smooth transition model. In this case we assume that a panel is available. The new model developed for this situation, the panel smooth transition regression [PSTR] model, is a non-dynamic panel model with individual fixed effects. It can be seen as an alternative to a linear model such that its coefficients are allowed to change smoothly as a function of an exogenous variable. The PSTR model can be useful in situations in which there is economic theory suggesting nonlinear relationships or when there are changing economic structures implying that the response parameters may vary over time and individuals according to a transition variable.

To be able to construct PSTR models in a systematic way, we extend the modelling strategy for smooth transition models introduced by Eitrheim and Teräsvirta (1996) and Teräsvirta (1998) to the panel framework. The strategy includes tests for parameter constancy and no remaining nonlinearity. Small-sample properties of these tests are investigated by simulation. The results indicate that the proposed tests perform in a satisfactory fashion in small samples. The PSTR model is used to investigate whether financial constraints affect firms' investment decisions. The results in this chapter support earlier findings by Hansen (1999) who, using a panel threshold model found that financial constrains do have an adverse effect on firms' investment behaviour.

Chapter 3. Simulation-based finite-sample linearity test against Smooth Transition models

In the third chapter we apply Monte Carlo testing techniques to the problem of testing linearity against smooth transition models. The Monte Carlo approach allows us to introduce a new test that differs from the tests existing in the literature in two respects. First, the test is exact in the sense that the probability of rejecting the null when it is true is always less that or equal to the nominal size of the test. Second, the test does not rely on an auxiliary regression obtained by replacing the model under the alternative by approximations based on a Taylor expansion as the test developed in Luukkonen, Saikkonen, and Teräsvirta (1988).

[2]This is a joint work with Timo Teräsvirta and Dick van Dijk.

The tests presented in this chapter can be used in situations in which the smooth transition model includes endogenous or exogenous variables in the explanatory variables. We also show how the Monte Carlo testing methods can be used for size-correcting the test of Luukkonen, Saikkonen, and Teräsvirta (1988). Small-sample properties of the test are investigated by simulation. The results indicate that our test not only has a correct size but that it also has good power properties. Furthermore, it becomes clear that the power loss implied by the Taylor approximation to the alternative model is not large when compared to the SupLM test. The power loss can be more substantial, however, when the ExpLM and AveLM are used. For definitions of these three statistics, see Andrews and Ploberger (1994).

In case of dynamic models we show that at sample sizes of 30 or 50 observations the test proposed by Luukkonen, Saikkonen, and Teräsvirta (1988) can be oversized when the null distribution is approximated by the asymptotic χ^2-distribution. The size distortion is large when the model under the alternative contains a large number of regressors as compared to the sample size and the model under the null is close to being nonstationary. In this case, Monte Carlo testing techniques can be very useful since they allow achieving size control while preserving the power.

Chapter 4. Smooth permanent surge process

Traditional models in time series analysis are either models in which all shocks are transitory and models in which they are not. For instance, in the stationary autoregressive model all shocks are transitory, whereas in the random walk all shocks are permanent. Recently, Engle and Smith (1999) proposed the stochastic permanent break [STOPBREAK] process to close the gap between transience and permanence. In the STOPBREAK model at one extreme, all shocks are transitory and at the other, they are permanent.

The most simple form of the STOPBREAK process is,

$$y_t = m_t + \epsilon_t \tag{2}$$

where ϵ_t is the innovation or shock, and m_t is a time-varying conditional mean defined by

$$m_t = m_{t-1} + q_{t-1}\epsilon_{t-1} \tag{3}$$

with $q_{t-1} = g(\epsilon_t)$. Function $g(\epsilon)$ is the logistic function

$$g(\epsilon_t) = \frac{\epsilon_t^2}{\gamma + \epsilon_t^2} \tag{4}$$

In the STOPBREAK model, the realized process at time t is a random walk when $q_{t-1} = 1$ in (3) and when $q_{t-1} = 0$, the conditional mean (3) does not change and neither does the long-run forecast for y_t. Consequently, in the STOPBREAK process the long-run effect of a shock depends on the value of q_{t-1}. Given the specification of $g(\epsilon)$ in (4) large shocks are permanent while small shock are only transitory. The main drawback of this specification is that only a zero shock has transitory effects, because (4) equals zero only when $\epsilon_t = 0$. Occurrence of a transitory shock is thus a zero probability event.

In this chapter we introduce a ne model called the smooth permanent surge [SPS] model as an alternative to the STOPBREAK process. In the SPS process, small shocks can be transitory while large shocks can be permanent. The SPS model is based on the moving average representation of (2) and (3) and has the following form:

$$\Delta y_t = \epsilon_t - \varpi_{t-1}\epsilon_{t-1} \tag{5}$$

where $\varpi_t = \theta_1 + \theta_2 g(\epsilon_t)$. Furthermore, $g(\epsilon)$ is the second-order logistic

$$g(\epsilon_t; \gamma, \mathbf{c}) = (1 + \exp(-\gamma(\epsilon_t - c_1)(\epsilon_t - c_2)))^{-1} \tag{6}$$

with $\gamma > 0$ and $c_1 \leq c_2$.

The process nests the STOPBREAK process and in the limiting case $\gamma \to \infty$, it converges to the threshold integrated moving average [TIMA] model by Gonzalo and Martínez (2003). The TIMA model is similar to (5) but ϖ_t is now an indicator function taking value θ_1 when $|\epsilon_t| \leq \kappa$ and θ_2 otherwise.

In the chapter we derive a test of SPS against the STOPBREAK process and introduce a new test for testing the SPS process against the random walk. Small-sample properties of these tests are investigated by Monte Carlo experiments. An application to the stock price series closes the chapter.

Bibliography

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

DUFOUR, J.-M., AND L. KHALAF (2001): "Monte Carlo test methods in econometrics," in *A companion to theoretical econometrics*, ed. by B. Baltagi, pp. 495–519, Oxford. Blackwell.

EITRHEIM, Ø., AND T. TERÄSVIRTA (1996): "Testing the adequacy of smooth transition autoregressive models," *Journal of Econometrics*, 74, 59–75.

ENGLE, R. F., AND A. D. SMITH (1999): "Stochastic permanent breaks," *The Review of Economics and Statistics*, 81, 553–574.

GONZALO, J., AND O. MARTÍNEZ (2003): "Threshold integrated moving average process (Does size matter? Maybe so)," *Preliminary draft.*

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

——— (1999): "Threshold effects in non-dynamic panels: Estimation, testing and Inference," *Journal of Econometrics*, 93, 345–368.

JANSEN, E., AND T. TERÄSVIRTA (1996): "Testing parameter constancy and super exogeneity in econometric equations," *Oxford Bulletin of Economics and Statistics*, 58, 735–763.

LIN, C. F., AND T. TERÄSVIRTA (1994): "Testing the constancy of regression parameters against continuous structural change," *Journal of Econometrics*, 62, 211–228.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing linearity against smooth transition autoregressive models," *Biometrics*, 75, 491–499.

TERÄSVIRTA, T. (1998): "Modelling economic relationships with smooth transition regressions," in *Handbook of applied economic statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. New York: Marcel Dekker.

VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): "Smooth transition autoregressive models - a survey of recent developments," *Econometric Reviews*, 21, 1–47.

Part II

The Chapters

Chapter 1

Testing parameter constancy in stationary vector autoregressive models against continuous change

1.1 Introduction

Parameter constancy or stability is one of the key assumptions in econometric models. The parameters of the models are estimated under this assumption, and it is also fundamental in forecasting with the estimated model. For a general discussion of parameter constancy in econometrics equations see Hendry (1996). Several stability tests are available for single-equation regression models, but the need for such tests is obvious in vector models as well. Lutkepohl (1991, pp. 159-166) discussed tests for the stationary vector autoregressive (VAR) model that are based on out-of-sample forecasts from the estimated VAR model. It is also possible, for example, to construct in-sample generalized Chow-tests for the VAR model as discussed by Andrews (1993) and obtain the critical values by applying the sampling techniques that Hansen (1996) suggested. Hansen and Johansen (1999) recently considered testing parameter constancy in cointegrated VAR models. Furthermore, Bai, Lumsdaine, and Stock (1998) derived asymptotically valid confidence intervals for a date of a break in a multivariate time series.

The above-mentioned approaches are, at least implicitly, based on the assumption that the alternative to constant parameters is a single structural break. While this is often a very sensible assumption, there may be occasions in which something more general could be needed. For example, model misspecification often shows not as a sharp break but as continuous change in parameters. In this chapter, we consider an in-sample parameter constancy test based on a parametric alternative to the null hypothesis. The alternative model allows the parameters to change smoothly over time. If the null hypothesis is rejected, the alternative model may be estimated. This may help to locate the misspecification and provide clues as to how to remove the causes of nonconstancy. Our alternative model is a special case of a vector smooth transition autoregressive (VSTAR) model that nests the linear constant-parameter VAR model. This means that we generalize the single-equation parameter constancy test of Lin and Teräsvirta (1994), see also Teräsvirta (1998), to the vector case. It should be mentioned that Ripatti and Saikkonen (2001) recently derived a test for testing the stability of the intercept in a cointegrating relationship within a nonstationary VAR model, using the same type of alternative as Lin and Teräsvirta (1994) did. Under the null hypothesis our VAR model is assumed stationary, and as in the single-equation case it turns out that standard asymptotic χ^2-based inference is applicable.

The plan of the chapter is as follows. In Section 1.2 we introduce the VSTAR model and its special case, the time varying VAR model. The test statistics are introduced in Section 1.3. The next section contains a simulation

study and our conclusions can be found in section 1.5. Proofs are in the appendix A.

1.2 The vector smooth transition autoregressive model

Consider the following pth order vector smooth transition autoregressive model

$$\mathbf{y}_t = \mu_0 + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{y}_{t-i} + \mathbf{G}(s_t)(\mu_1 + \sum_{i=1}^{p} \mathbf{\Theta}_i \mathbf{y}_{t-i}) + \varepsilon_t, \ t = 1, \ldots, T \qquad (1.1)$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$ is an $(m \times 1)$ random vector, μ_i, $i = 0, 1$, are $(m \times 1)$ constant parameter vectors, $\mathbf{\Phi}_i$ and $\mathbf{\Theta}_i$, $i = 1, \ldots, p$, are $(m \times m)$ parameter matrices, and ε_t is an $(m \times 1)$ error vector. The $(m \times m)$ diagonal matrix of transition functions $\mathbf{G}(t) = \text{diag}\{G_1(s_t, \gamma_1; c_1), \ldots, G_m(s_t, \gamma_m; c_m)\}$ allows the model to change smoothly from one extreme regime to the other as a function of the transition variable s_t. The diagonal elements $G_i(s_t, \gamma_i; c_i)$, $i = 1, \ldots, m$, are transition functions such that $0 \leq G_i(s_t, \gamma_i; c_i) \leq 1$.

More specifically, we define the transition functions as follows:

$$\begin{aligned} G_i(s_t) &= G_i(s_t; \gamma_i; c_{i_j}) \\ &= (1 + \exp\{-\gamma_i \prod_{j=1}^{k} (s_t - c_{i_j})\})^{-1} \end{aligned} \qquad (1.2)$$

where $\gamma_i > 0$ and $c_{i_1} \leq c_{i_2} \leq \cdots \leq c_{i_k}$ for $i = 1, \ldots, m$. These parameters restrictions are identifying restrictions.

Equations (1.1) and (1.2) form a logistic vector smooth transition autoregressive (LVSTAR) model. In the chapter, we assume $s_t = t$ and call this special case the time-varying vector autoregressive (TV-VAR) model. It is our maintained model in testing parameter constancy of a linear vector autoregressive model. The TV-STAR model is parametric, and this allows us to develop parametric tests against parameter instability.

Some special cases of the TV-STAR model are of interest. When $k = 1$ and $c_1 = \cdots = c_m = c$ and, furthermore, we let $\gamma_i \to \infty, i = 1, \ldots, m$, in $\mathbf{G}(t)$, the TV-STAR model becomes a VAR model with a single structural break at $\frac{t}{T} = c$. When $\mathbf{\Theta}_i = \mathbf{0}, i = 1, \ldots, p$, the model collapses into an important special case in which only the intercept changes over time whereas the dynamic structure of the model does not. Setting $\gamma_i = 0, i = 1, \ldots, m$, makes the model a linear VAR model. When $m = 1$, the TV-STAR model becomes an LSTAR model with time as the transition variable, considered

in Lin and Teräsvirta (1994). Clearly, (1.1) and (1.2) form a very flexible parametrization capable of characterizing many different types of parameter nonconstancy, from slow, very smooth change in parameters to very rapid changes, including breaks. This includes situations in which the change occurs at either end of the observation period and ones in which the parameters are constant in the beginning and end but there is a change somewhere in the middle.

1.3 Testing parameter constancy in the linear VAR model

Our preferred null hypothesis in testing parameter constancy of a linear VAR model against TV-VAR is $H_0 : \gamma_i = 0, i = 1, \ldots, m$. Note, however, that (1.1) may be made linear also by setting $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\Theta}_i = \mathbf{0}$, $i = 1, \ldots, p$, in (1.1). This is an indication of the fact that (1.1) is not identified under the null hypothesis. When H_0 holds, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Theta}_i$, $i = 1, \ldots, p$, and $\mathbf{c}_i$, $i = 1, \ldots, m$, are not identified and may thus obtain any value without the value of the criterion function being affected. Conversely, if $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\Theta}_i = \mathbf{0}$, $i = 1, \ldots, p$, then γ_i and $\mathbf{c}_i$, $i = 1, \ldots, m$, are not identified. This has the consequence that the standard asymptotic inference is not available as the asymptotic distribution of the likelihood ratio statistic under the null hypothesis is not known. For discussion see Davies (1977,1987) and Hansen (1996).

In this chapter, we circumvent this difficulty by an approximation of the diagonal elements (1.2) of the transition matrix $\mathbf{G}(t)$. For notational convenience, we replace $G(t; \gamma_i; \mathbf{c}_i)$ by $\widetilde{G}(t; \gamma_i; \mathbf{c}_i) = G(t; \gamma_i, \mathbf{c}_i) - G(t; 0, \mathbf{c}_i)$ for $i = 1, \ldots, m$. This results in $\widetilde{G}(t; 0, \mathbf{c}_i) = 0$, but the transformation does not affect the generality of the argument. This approach is a generalization of the one in Lin and Teräsvirta (1994), see also Luukkonen, Saikkonen, and Teräsvirta (1988), to the vector case.

We make the following assumptions about model (1.1) under H_0 :

(A1) The stochastic sequence $\{\boldsymbol{\varepsilon}_t\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\mathcal{F}_t$ with

$$\sup_t \mathsf{E}\{|\varepsilon_{jt}|^{2+\alpha} | \mathcal{F}_{t-1}\} < \infty \quad \text{a.s.}$$

for some $\alpha > 0$ and $j = 1, \ldots, m$, and

$$\lim_{t \to \infty} \mathsf{E}\{\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}\} = \boldsymbol{\Sigma}_\varepsilon \quad \text{a.s.}$$

for some positive definite matrix $\boldsymbol{\Sigma}_\varepsilon$.

(A2) The VAR process is stable, that is, $|\mathbf{I} - \sum_{j=1}^{p} \mathbf{\Phi}_j z^j| \neq 0, |z| \leq 1$, and $\mathsf{E}|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| < \infty$ for all i, j, k, l and t.

These assumptions guarantee the existence of the second moments for $\mathbf{y}_t$ and the convergence of the sample moments to their true values. Assumption (A1) implies $\mathsf{E}\varepsilon_t\varepsilon_t' = \mathbf{\Sigma}_\varepsilon$.

Consider now the first-order Taylor approximation to $\widetilde{G}(t; \gamma_i, \mathbf{c}_i)$ about $\gamma_i = 0$:

$$\begin{aligned}\widetilde{G}(t; \gamma_i, \mathbf{c}_i) &= \gamma_i \frac{\partial}{\partial \gamma_i} \widetilde{G}(t; 0, \mathbf{c}_i) + R_i \\ &= \alpha_i + (1/4)\gamma_i \prod_{j=1}^{k} (t - c_{ij}) + R_i \quad i = 1, \ldots, m, \end{aligned} \tag{1.3}$$

where R_i is the remainder. Substituting (1.3) for $G(t; \gamma_i, \mathbf{c}_i)$, $i = 1, \ldots, m$, in (1.1) and rearranging terms yields

$$\mathbf{y}_t = \mathbf{B}_0\mathbf{w}_t + \mathbf{B}_1\mathbf{w}_t t + \cdots + \mathbf{B}_k\mathbf{w}_t t^k + \varepsilon_t^* \tag{1.4}$$

where $\mathbf{B}_0$ is an $m \times (mp + 1)$ coefficient matrix, $\mathbf{B}_i = \mathbf{\Gamma}\mathbf{B}_i^*$, $i = 1, \ldots, k$, are $m \times (mp + 1)$ coefficient matrices such that $\mathbf{B}_i^* \neq \mathbf{0}$, $i = 1, \ldots, k$, and $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \ldots, \gamma_m)$. Furthermore, $\mathbf{w}_t = (1, y_{1,t-1}, \ldots, y_{1,t-p}, \ldots, y_{m,t-1}, \ldots, y_{m,t-p})'$ is an $(mp + 1) \times 1$ vector and

$$\varepsilon_t^* = \mathbf{R}(\mu_1 + \sum_{i=1}^{p} \mathbf{\Theta}_i \mathbf{y}_{t-i}) + \varepsilon_t$$

where the remainder matrix $\mathbf{R} = \mathrm{diag}(R_1, \ldots, R_m)$. Note that under H_0, $\varepsilon_t^* = \varepsilon_t$ so that the distributional properties of the error process are not affected by the approximation when the null hypothesis is valid. Also note that under H_0, $\mathbf{B}_0 = [\mu_0, \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p]$.

It is useful to rewrite (1.4) as follows:

$$\mathbf{y}_t = \mathbf{B}\mathbf{z}_t + \varepsilon_t^* \tag{1.5}$$

where $\mathbf{z}_t = \mathbf{s}_t \otimes \mathbf{w}_t$ with $\mathbf{s}_t = (1, \tilde{\mathbf{s}}_t')'$, $\tilde{\mathbf{s}}_t = \left(t, \ldots, t^k\right)'$ and $\mathbf{B} = (\mathbf{B}_0, \mathbf{B}_1^*, \ldots, \mathbf{B}_k^*)$ is an $(m \times ((k + 1)(mp + 1)))$ parameter matrix.

The idea of the approximation is now obvious. As $\mathbf{B}_i = \mathbf{\Gamma}\mathbf{B}_i^*$ and $\mathbf{B}_i^* \neq \mathbf{0}$, $i = 1, \ldots, k$, the original null hypothesis $\mathbf{\Gamma} = \mathbf{0}$ is replaced by the hypothesis

$$\mathrm{H}_0' : \mathbf{B}_i = \mathbf{0}, i = 1, \ldots, k \tag{1.6}$$

which is a linear hypothesis within a linear vector autoregressive model. We have traded off information about the structure of the alternative model in order to obtain a simple testing situation.

In order to carry out the test, the parameters of (1.4) have to be estimated. The quasi maximum likelihood estimator of $\mathbf{B}$ can be written as

$$\mathbf{b}_T=\text{vec}(\widehat{\mathbf{B}}_T) = ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z}\otimes\mathbf{I}_m)\text{vec}(\mathbf{Y}) \tag{1.7}$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$ and $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T)$, see, for example, Lutkepohl (1991, p. 63). Under (A1) and (A2) it can be shown that $\mathbf{b}_T$ is asymptotically normal. The result is given in the following theorem:

Theorem 1 *Consider the auxiliary regression (1.5) and assume that under (1.6), assumptions (A1) and (A2) hold. Then*

$$(\mathbf{\Upsilon}_T\otimes\mathbf{I}_m)(\mathbf{b}_T-\beta) \xrightarrow{d} \mathsf{N}(\mathbf{0}, \mathbf{\Lambda}^{-1}\otimes\mathbf{\Sigma}_\varepsilon) \tag{1.8}$$

as $T\to\infty$, *where* $\beta = vec(\mathbf{B})$, $\mathbf{\Upsilon}_T = \mathbf{D}_k\otimes\mathbf{I}_{mp+1}$, *and* $\mathbf{D}_k = diag(T^{1/2}, T^{3/2}, \ldots, T^{(2k+1)/2})$ *such that* $\mathbf{\Lambda} = p\lim_{T\to\infty}\mathbf{\Upsilon}_T^{-1}(\mathbf{Z}\mathbf{Z}')\mathbf{\Upsilon}_T^{-1}$.

Proof. See Appendix A. ■

This result makes it possible to make use of standard asymptotic distribution theory in testing the null hypothesis (1.6). We can also test subhypotheses assuming that a subset of parameters are constant under the alternative. For example, if we want to test the hypothesis that the intercepts are constant, equation (1.4) has the form

$$\mathbf{y}_t= \mathbf{B}_0\mathbf{w}_t+\mathbf{b}_1 t + \cdots + \mathbf{b}_k t^k + \varepsilon_t^*$$

where $\mathbf{b}_i, i = 1, \ldots, k$, are $(m\times 1)$ parameter vectors. We can use the analogue likelihood ratio (LR) type statistic, the Wald type statistic and the Lagrange multiplier type statistic for testing the null hypothesis. In the general case, these statistics have an asymptotic χ^2 distribution with $km(mp+1)$ degrees of freedom when the null hypothesis holds.

The performance of the tests depends on the choice of parameter k. If the parameters (or a subset under test) are assumed to change monotonically, $k = 1$ suffices, because the change can in that case be characterized by a simple logistic function obtained by setting $k = 1$ in (1.2). Setting $k = 2$ is enough when one wants to include the case where the parameters first change in one direction and then start changing back toward their original values. In this case, the test based on $k = 1$ does not usually have power. The choice $k = 3$ allows for more complicated, possibly nonmonotonic, change. In practice it

is advisable to carry out the test at least for $k = 1, 2$. If the time series are sufficiently long, choosing $k = 3$ may also be an option. Comparing the strengths of the rejection (p-values) often gives preliminary information about the type of the parameter change that has taken place. The point of defining parameter change by equation (1.2) is that it allows defining changes that both begin and end well within the observation period. A single break serves as an extreme example. This is not possible if the change is fully characterized by a polynomial of t. For this reason equation (1.4) is only an auxiliary regression that is not assumed to generate any data. But then, the test that we propose naturally does have power against the less likely alternative that the change in parameters is accurately described by a polynomial of time.

In order to define these three statistics, let $\widetilde{\mathbf{E}}_R = (\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_T)'$ and $\widehat{\mathbf{E}}_U = (\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_T)'$ where $\widetilde{\varepsilon}_t$ and $\widehat{\varepsilon}_t$, $t = 1, \ldots, T$, are the restricted and unrestricted estimated residuals, respectively, from (1.4). We define the following cross product matrices: $\mathbf{S}_R = \widetilde{\mathbf{E}}_R'\widetilde{\mathbf{E}}_R$, $S_U = \widehat{\mathbf{E}}_U'\widehat{\mathbf{E}}_U$, and furthermore, set $\mathbf{W}_1 = (\mathbf{w}_1, \ldots, \mathbf{w}_T)'$ and $\mathbf{W}_2 = \{(\tilde{\mathbf{s}}_1 \otimes \mathbf{w}_1), (\tilde{\mathbf{s}}_2 \otimes \mathbf{w}_2), \ldots, (\tilde{\mathbf{s}}_T \otimes \mathbf{w}_T)\}'$. Then the Wald, LM and LR test statistics can be written in the following way:

$$
\begin{aligned}
W &= T(\mathrm{tr}\mathbf{S}_U^{-1}\mathbf{S}_R - m) \\
LM &= T\mathrm{tr}\left(\mathbf{S}_R^{-1}\widetilde{\mathbf{E}}_R'\mathbf{W}_2\left[\mathbf{W}_2'\mathbf{W}_2 - \mathbf{W}_2'\mathbf{W}_1\left(\mathbf{W}_1'\mathbf{W}_1\right)^{-1}\mathbf{W}_1'\mathbf{W}_2\right]^{-1}\mathbf{W}_2'\widetilde{\mathbf{E}}_R\right) \\
LR &= T\ln\left[\frac{\det\left(\mathbf{S}_R\right)}{\det\left(\mathbf{S}_U\right)}\right].
\end{aligned}
$$

In the simulations we also use the F-versions of these tests:

$$
F_W = \frac{T-(k+1)(1+pm)}{k(1+pm)mT}W \overset{\text{approx}}{\sim} F\left[k(1+pm)m, T-(k+1)(1+pm)\right] \tag{1.9}
$$

$$
F_{LR} = \frac{T-(1+pm)}{k(1+pm)mT}LR \overset{\text{approx}}{\sim} F\left[k(1+pm)m, T-(1+pm)\right] \tag{1.10}
$$

$$
F_{LM} = \frac{1}{k(1+pm)m}LM \overset{\text{approx}}{\sim} F\left[k(1+pm)m, T\right] \tag{1.11}
$$

under H_0. The reason for this is that while they are not exact F-test, their small-sample properties can be expected to be better than their asymptotic counterparts. The simulation study and its results will be discussed in the next section.

1.4 Simulation study

1.4.1 Rao's test and design of the experiment

A recurring problem of tests of VAR models is that the standard test statistics tend to be oversized in small samples. This may even be true for the F-versions (1.9), (1.10) and (1.11). Rao (1973, p. 556) suggested yet another F-statistic to remedy this problem. It is based on the likelihood ratio LR and defined as follows:

$$F_{\text{RAO}} = \left(\frac{1 - LR^{-1/s}}{LR^{-1/s}} \right) \frac{\delta s - 2\lambda}{w} \tag{1.12}$$

where $w = mk(pm+1)$, $s^2 = \left(w^2 - 4\right) / \left(m^2 + r^2 - 5\right)$, $r = k(1+pm)$, $\delta = T - 0.5(m+r+1)$ and $\lambda = 0.25(w-2)$. We include it in our simulations.

Another solution to the potential size distortion is to generate empirical distributions for the test statistic by parametric bootstrap and use them instead of the χ^2 and F-distributions discussed above. In order to do that, we assume that the errors in (1.4) are independent and identically distributed under the null hypothesis. We can then apply the recursive bootstrap, see Li and Maddala (1996), as follows:

1. Estimate the parameters of (1.4) under the null hypothesis and compute the matrix of residuals $\widetilde{\mathbf{E}}_R$.

2. Generate N matrices $\widetilde{\mathbf{E}}_{R1}, \ldots, \widetilde{\mathbf{E}}_{RN}$ by sampling vectors $\widetilde{\varepsilon}_t$ from $\widetilde{\mathbf{E}}_R$ with replacement.

3. Given these matrices, compute N realizations of size T of $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$, call them $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(N)}$, recursively, using the estimated values of the null model (no data are generated by the auxiliary regression) and the first $p+1$ observations of the original sample, $\mathbf{y}_{-p}, \ldots, \mathbf{y}_0$, as starting values.

4. Estimate the auxiliary model (1.4) under the null hypothesis and the alternative for each data set $\mathbf{Y}^{(i)}, i = 1, \ldots, N$ and compute the value of the test statistics for each realization.

The empirical p-value of the test is obtained as the fraction of times the value of the bootstrap statistic exceeds the values obtained from the original sample[1]

[1] In this case, the boostrap method can be justified with asymptotic arguments since the statistics considered are smooth functions of sample moments, and they are also asymptotically pivotal [see Horowitz (2001)].

The design of the experiment is the following. The sample sizes are $T = 50$ and 100. The VAR models have either one or two lags, and the dimension of the model varies between two and four equations. The power of t in the artificial VAR models, called k, see equation (1.4), equals either 1 or 2. The definitions of the six models used in the simulation experiment can be found in Appendix B. Furthermore, the number of replications in every individual experiment is 1000 and the nominal significance level of the tests equals 0.05. The bootstrap p-value is computed using 1000 bootstrap samples of equal size as the original sample.

1.4.2 Size simulations

The size simulations involve the models called DGP1, DGP 2, DGP 3 and DGP 4 in Appendix B. In size simulations, the models have been simulated without their nonlinear component, that is, assuming $\mu_1 = \mathbf{0}$ and $\Theta_1 = \Theta_2 = 0$ in (1.1). The results can be found in Table 1.1. It is seen that the size of the asymptotic tests can be badly biased, especially when the number of parameters under the null hypothesis is large compared to the sample size; see the worst case $T = 50$, $m = 4$, $p = 2$ and $k = 2$. This is not unexpected because we are testing a linear hypothesis within a linear (auxiliary) system; for comparable results see, for instance Edgerton and Shukur (1999) and Candelon and Lütkepohl (2001). The Wald type tests perform particularly badly, erroneously rejecting the null hypothesis in 100% of the cases in this set-up. This happens because it relies solely on the estimation of the alternative (larger) model. The LM type test performs much better but not sufficiently well either. Even Rao's F-statistic (1.12) is generally somewhat oversized when $T = 50$, but the situation improves for $T = 100$. It is seen that the F-test represent a considerable improvement over the χ^2-tests, but their likelihood ratio and Wald versions can still be oversized for $T = 50$. Perhaps surprisingly, the F-LM type test can be considerably undersized in small samples. This happens in particular when $k = 2$.

Finally, the lower panel of Table 1.1 contains the empirical sizes of the bootstrapped versions of these statistics. The χ^2 and the F version of the tests are equivalent, and because of that p-values of the latter ones are omitted from the table. This outcome is due to the fact that with the empirical distribution of the test, the small sample correction evident in (1.9), (1.10) and (1.11) does not play a role. Two main results emerge. First, the bootstrap improves the empirical size and, second, the differences between the sizes of the tests practically vanish. This suggests that the LM-type test is a natural candidate to be bootstrapped because it requires fewer computations than the other

tests. This is due to the fact that the auxiliary VAR equation only needs to be estimated under the null hypothesis.

1.4.3 Power simulations

The power properties of our tests are investigated using eight different model specifications. The DGPs are again DGP 1, DGP 2, DGP 3 and DGP4 in Appendix B but now with their nonlinear components included. Each parameter combination serves in two different TV-VAR models, as both the first-order ($k = 1$) and the second-order ($k = 2$) logistic model have been simulated. In order to avoid size distortions we only apply bootstrap-based tests here. The transition function is the same for all equations, with the slope parameter $\gamma = 10$ or 20, the location parameter $c_1 = (0.5)\,T$ for $k = 1$, and $c_1 = (0.3)\,T$ and $c_2 = (0.7)\,T$ for $k = 2$.

The results for $k = 1$ can be found in Table 1.2. It can be seen that the differences in power are generally not large, which supports our previous assertion that among the bootstrapped tests, the LM type test would be the one to favour because its use would minimize the amount of computations. The results indicate that the power of the test increases with the value of γ, the slope parameter. This is what intuition might suggest, that is, a break or near-break is easier to detect than a very smooth transition, *ceteris paribus*. The results for $k = 2$ have a similar pattern, and for this reason they are not reported here. In order to get a better idea of the relative power of our tests we compare them to the generalized Chow-test with an unknown break-point that Andrews (1993) considered. In addition to the supremum test of Andrews we follow Andrews and Ploberger (1994) and also include the "average test" statistic

$$\text{ave}S_T(\pi) = \int_{\pi\in\Pi} S_T(\pi)dW(\pi)$$

and the "exponential average test" statistic

$$\exp S_T(\pi) = \ln(\int_{\pi\in\Pi} \exp\{(1/2)S_T(\pi)\}dW(\pi)).$$

where the weight function $W(\pi)$ allocates equal weights on all observations πT such that $\pi \in \Pi$. Furthermore, $S_T(\pi)$ is the Chow-statistic assuming a break at πT . We use the LM-versions of these tests and calculate the critical values of the test statistics following the procedure of Hansen (1996). We use 500 draws from a normal distribution to compute the asymptotic p-values and 500 bootstrap samples for computing the empirical p-values. We report

Table 1.1: Empirical sizes of the parameter constancy tests based on asymptotic and bootstrapped p-values

Results based on asymptotic or approximate p-values								
				k=1				
	(DGP 1)		(DGP 3)		(DGP 2)		(DGP 4)	
	(m=2,p=1)		(m=4,p=1)		(m=2,p=2)		(m=4,p=2)	
Statistic	T=50	T=100	T=50	T=100	T=50	T=100	T=50	T=100
W	0.132	0.081	0.357	0.156	0.214	0.108	0.810	0.344
LR	0.106	0.072	0.220	0.108	0.145	0.083	0.556	0.205
LM	0.072	0.063	0.089	0.055	0.084	0.061	0.198	0.085
F_{RAO}	0.058	0.062	0.054	0.049	0.051	0.050	0.048	0.046
F_W	0.061	0.062	0.068	0.051	0.060	0.050	0.057	0.041
F_{LR}	0.051	0.059	0.044	0.043	0.047	0.049	0.054	0.040
F_{LM}	0.035	0.053	0.021	0.037	0.037	0.044	0.041	0.033
				k=2				
W	0.229	0.118	0.766	0.307	0.488	0.206	1.000	0.801
LR	0.143	0.076	0.482	0.170	0.267	0.131	0.973	0.473
LM	0.061	0.051	0.133	0.070	0.095	0.070	0.399	0.136
F_{RAO}	0.061	0.048	0.068	0.058	0.063	0.058	0.058	0.062
F_W	0.062	0.049	0.076	0.055	0.060	0.057	0.077	0.045
F_{LR}	0.055	0.048	0.084	0.057	0.081	0.060	0.297	0.092
F_{LM}	0.030	0.040	0.014	0.034	0.025	0.046	0.008	0.025
Results based on bootstrapped p-values								
				k=1				
	(DGP 1)		(DGP 3)		(DGP 2)		(DGP 4)	
	(m=2,p=1)		(m=4,p=1)		(m=2,p=2)		(m=4,p=2)	
Statistic	T=50	T=100	T=50	T=100	T=50	T=100	T=50	T=100
W	0.054	0.056	0.049	0.050	0.053	0.051	0.052	0.048
LR	0.053	0.056	0.046	0.046	0.052	0.053	0.052	0.045
LM	0.052	0.057	0.054	0.045	0.055	0.054	0.046	0.040
F_{RAO}	0.053	0.056	0.046	0.046	0.052	0.053	0.052	0.045
				k=2				
W	0.049	0.045	0.058	0.051	0.043	0.056	0.052	0.059
LR	0.048	0.046	0.053	0.053	0.047	0.055	0.056	0.058
LM	0.048	0.044	0.056	0.053	0.048	0.054	0.064	0.052
F_{RAO}	0.048	0.046	0.053	0.053	0.047	0.055	0.056	0.058

Table 1.2: Power of the bootstrapped parameter constancy tests

		k=1							
		(DGP 1)		(DGP 3)		(DGP 2)		(DGP 4)	
		(m=2,p=1)		(m=4,p=1)		(m=2,p=2)		(m=4,p=2)	
	Statistic	T=50	T=100	T=50	T=100	T=50	T=100	T=50	T=100
$\gamma = 10$	W	0.432	0.887	0.706	1.000	0.459	0.910	0.547	0.989
	LR	0.431	0.887	0.693	1.000	0.449	0.905	0.543	0.987
	LM	0.427	0.887	0.667	0.998	0.444	0.902	0.505	0.982
	F_{RAO}	0.431	0.887	0.693	1.000	0.449	0.905	0.543	0.987
$\gamma = 20$	W	0.485	0.903	0.625	0.977	0.505	0.951	0.561	0.997
	LR	0.478	0.901	0.629	0.978	0.502	0.950	0.557	0.997
	LM	0.468	0.899	0.621	0.976	0.488	0.949	0.515	0.996
	F_{RAO}	0.478	0.901	0.629	0.978	0.502	0.950	0.557	0.997

results from two different simulations. The number of replications equals 1000 as before, and the sample sizes are 50 and 100.

In the first simulation the model is DGP 1b in Appendix B, and the observations are generated assuming that $k = 1$ and, furthermore, that the parameters change smoothly over time such that $c = 0.5T$ and $\gamma = 0, 5, 10$. The first value $\gamma = 0$ refers to the size simulation. The results of this experiment are given in Table 1.3. It is seen that the generalized Chow-test with asymptotic critical values is undersized in all three cases when $T = 50$, whereas the same test based on bootstrapped critical values does not suffer from this disadvantage. As to our test, their bootstrapped versions have similar power, and for this reason we only report the results of the LM test. Comparing the Chow-tests with the LM test, it is seen that at both samples sizes our tests have better empirical power than the Chow-tests for $\gamma = 5$ and $\gamma = 10$. This is not unexpected, because the change in parameters is smooth.

In the second experiment, the model, DGP 2b in Appendix B, has a single structural break halfway through the observation period. The power function is estimated for three different cases, consisting of the null model and two different sets of parameter values under the alternative. The parameters under the alternative are defined by subtracting a fixed number from all parameters in the null model. The difference between the parameters of the null model is called "increment" in the table.

The results can be found in Table 1.4. They show that our test also has power against an abrupt change. However, in this case the generalized Chow-test must have higher power than our test statistics do, because the Chow-test is a test explicitly designed against a single structural break. In our smooth transition-based tests, the alternative is larger and only contains the true

Table 1.3: Size and power of the recursive Chow-test and the bootstrapped parameter constancy test when the alternative is a smooth change

		Asymptotic		Bootstrap	
Gamma	Statistic	T=50	T=100	T=50	T=100
	LM			0.055	0.053
$\gamma = 0$	SupLM	0.029	0.050	0.060	0.043
	ExpLM	0.030	0.048	0.058	0.049
	AveLM	0.041	0.050	0.050	0.045
	LM			0.169	0.362
$\gamma = 5$	SupLM	0.085	0.323	0.121	0.279
	ExpLM	0.087	0.333	0.133	0.286
	AveLM	0.122	0.374	0.162	0.334
	LM			0.274	0.611
$\gamma = 10$	SupLM	0.159	0.584	0.198	0.488
	ExpLM	0.179	0.594	0.228	0.528
	AveLM	0.227	0.657	0.257	0.604

alternative as a special case. This causes a power loss when compared to the Chow-tests where the alternative is just a single structural break. This is also what the results in Table 1.4 indicate. The power of the Chow-tests increases faster than that of our tests when the sample size is doubled from 50 to 100, and the difference in power is quite clear at the larger sample size. Finally, we investigate the power of the LM test when the alternative model is the TV-VAR model with $k = 2$ in (1.2) and the value of k in the auxiliary regression (1.4) is either $k = 1$ or 2. For comparison, we include the recursive Chow-test in the experiment. The simulations are based on DGP 1b. The results in Table 1.5 indicate that when the alternative is a TV-VAR model with $k = 2$ in (1.2) both the recursive Chow test and the LM test using $k = 1$ in (1.4) have low power. Not surprisingly, the power of the LM test when $k = 2$ in the auxiliary regression is much higher than the power of the recursive Chow test.

Two conclusions emerge from this experiment. First, the recursive Chow test is not powerful when there are two parameters changes of opposite sign and size in the sample. Second, if one is interested in estimating the alternative model (1.1) and (1.2) when the null hypothesis is rejected. The following testing sequence for selecting between $k = 1$ and $k = 2$ in (1.2) may be recommended. Test parameter constancy using (1.4) with $k = 2$. If the null is rejected, test the same null hypothesis using $k = 1$ in (1.4). Select $k = 1$ if

Table 1.4: Size and power of the recursive Chow test and the bootstrapped parameter constancy test when the alternative is an abrupt change

		Asymptotic		Bootstrap	
Increment	Statistic	T=50	T=100	T=50	T=100
	LM			0.054	0.052
0.0	SupLM	0.059	0.063	0.055	0.043
	ExpLM	0.048	0.057	0.054	0.050
	AveLM	0.068	0.068	0.060	0.055
	LM			0.145	0.330
-0.15	SupLM	0.157	0.495	0.163	0.431
	ExpLM	0.146	0.471	0.139	0.406
	AveLM	0.155	0.449	0.126	0.370
	LM			0.290	0.655
-0.30	SupLM	0.468	0.927	0.467	0.909
	ExpLM	0.435	0.922	0.432	0.905
	AveLM	0.382	0.868	0.306	0.818

the latter test rejects the null hypothesis otherwise select $k = 2$.

Table 1.5: Power of the recursive Chow test and the bootstrapped parameter constancy test when the alternative is a smooth change with $k = 2$.

		Asymptotic		Bootstrap	
Gamma	Statistic	T=50	T= 100	T=50	T=100
	LM($k = 1$)	0.039	0.045	0.038	0.044
	LM($k = 2$)	0.185	0.516	0.193	0.526
$\gamma = 5$	SupLM	0.031	0.108	0.079	0.174
	ExpLM	0.028	0.098	0.075	0.152
	AveLM	0.028	0.070	0.064	0.110
	LM($k = 1$)	0.042	0.051	0.040	0.049
	LM($k = 2$)	0.430	0.902	0.434	0.904
$\gamma = 10$	SupLM	0.048	0.304	0.129	0.406
	ExpLM	0.049	0.279	0.129	0.384
	AveLM	0.049	0.167	0.100	0.237

Note: LM($k = 1$) and LM($k = 2$) denote the LM test based on auxiliary regression (4) with $k = 1$ and $k = 2$, respectively.

1.5 Conclusions

We have derived a test of parameter constancy for a stationary vector autoregressive model, when the alternative consists of smoothly changing parameters. As previous studies have already suggested, asymptotic distribution theory does not work well here, and the use of empirical null distributions obtained

by a recursive bootstrap appears necessary in order to avoid size distortion. An important advantage of our tests is that they are computationally easy to carry out. Even when the alternative model is nonlinear, only linear VAR models need to be estimated and simulated. Nevertheless, in some situations where the null hypothesis is eventually rejected it may be of interest to try and estimate the corresponding TV-VAR model. If this is successfully done, the estimated model is likely to contain useful information about the nature of parameter change. Such information may in turn be helpful in finding ways to respecify the model such that the new model does have constant parameters. Our tests may thus be best viewed as easily applicable model-building tools for users of VAR models.

Bibliography

ANDREWS, D. W. K. (1993): "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, 821–856.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

BAI, J., R. L. LUMSDAINE, AND J. H. STOCK (1998): "Testing for and dating common breaks in multivariate time series," *Review of Economic Studies*, 65, 395–432.

BROWN, B. M. (1971): "Martingale central limit theorems," *Annals of Mathematical Statistics*, 42, 59–66.

CANDELON, B., AND H. LÜTKEPOHL (2001): "On the reliability of the Chow type test for parameter constancy in multivariate dynamic models," *Economics Letters*, 73, 155–160.

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 alternative.," *Biometrika*, 74, 33–43.

DVORETZKY, A. (1972): "Asymptotic normality for sums of dependent random variables," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 513–535, Berkeley. University of California Press.

EDGERTON, D., AND G. SHUKUR (1999): "Testing autocorrelation in a system perspective," *Econometric Reviews*, 18, 343–386.

HALL, P., AND C. C. HEYDE (1980): *Martingale limit theory and its application.* Academic Press, New York and London.

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

HANSEN, H., AND S. JOHANSEN (1999): "Some tests for parameter constancy in cointegrated VAR-models," *Econometrics Journal*, 2, 306–333.

HENDRY, D. F. (1996): "On the constancy of time-series econometrics equations," *Economic and Social Review*, 27, 401–422.

HOROWITZ, J. L. (2001): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 5, Amsterdam. Elsevier.

LAI, T. L., AND H. ROBBINS (1981): "Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes.," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 56, 329–360.

LAI, T. L., AND C. Z. WEI (1982): "Least squares estimates in stochastic regression models with applications in identification and control of dynamic Systems," *Annals of Statistics*, 10, 154–166.

LI, H., AND G. S. MADDALA (1996): "Bootstrapping time series models," *Econometric Reviews*, 15, 115–158.

LIN, C. F., AND T. TERÄSVIRTA (1994): "Testing the constancy of regression parameters against continuous structural change," *Journal of Econometrics*, 62, 211–228.

LÜTKEPOHL, H. (1991): *Introduction to multiple time series analysis.* New York: Springer-Verlag, 2 ed.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

RAO, C. R. (1973): *Linear statistical inference and its applications.* New York: Wiley, 2nd ed.

RIPATTI, A., AND P. SAIKKONEN (2001): "Vector autoregressive processes with nonlinear time trends in cointegrating relations," *Macroeconomic Dynamics*, 5, 577–597.

TERÄSVIRTA, T. (1998): "Modelling economic relationships with smooth transition regressions," in *Handbook of applied economic statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. New York: Marcel Dekker.

VAN DIJK, D. (1999): *Smooth transition models: extensions and outlier robust inference*. Tinbergen Institute, Erasmus University Rotterdam, PhD Thesis.

Appendix A

Proof of theorem

Consider the vector regression model

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{x}_{ti} + \varepsilon_t = \mathbf{B}\mathbf{z}_t + \varepsilon_t, \ t = 1, \cdots, T \tag{A.1}$$

where $\mathbf{y}_t = (y_{1t}, \cdots, y_{mt})'$ is an $(m \times 1)$ random vector, $\mathbf{A}_i$, $i = 1, \cdots, p$, are $(m \times m)$ parameter matrices, $\mathbf{B} = (\mathbf{A}_1, ..., \mathbf{A}_p)$ is an $(m \times mp)$ matrix, $\mathbf{x}_{ti} = (x_{1t,i}, \cdots, x_{mt,i})'$, $i = 1, ..., p$, are random vectors, $\mathbf{z}_t = (\mathbf{x}'_{t1}, ..., \mathbf{x}'_{tp})'$ is an $(mp \times 1)$ vector and ε_t is an $(m \times 1)$ error vector. Rewrite (A.1) as

$$\mathbf{y} = (\mathbf{Z}' \otimes \mathbf{I}_m)\beta + \varepsilon$$

where $\beta = vec(\mathbf{B})$, $\mathbf{y} = (\mathbf{y}'_1, ..., \mathbf{y}'_T)'$ is an $(mT \times 1)$ vector, $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_T)$ is an $(mp \times T)$ matrix and, furthermore, $\varepsilon = (\varepsilon'_1, ..., \varepsilon'_T)'$ is an $(mT \times 1)$ error vector. We first prove the consistency and asymptotic normality of the least squares estimator of β. The following lemma is a generalization of Theorem 3 of Lai and Wei (1982) to the vector case.

Lemma Consider the $(m \times 1)$ vector regression model (A.1) and assume that:

(a) The stochastic sequence $\{\varepsilon_t\}$ in (A.1) is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$ such that

$$\sup_t \mathsf{E}\{|\varepsilon_{jt}|^{2+\alpha} | \mathcal{F}_{t-1}\} < \infty \ \text{ a.s.} \tag{a1}$$

for some $\alpha > 0$ and $j = 1, ..., m$, and

$$\lim_{t \to \infty} \mathsf{E}\{\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1}\} = \mathbf{\Sigma}_\varepsilon \ \text{ a.s.} \tag{a2}$$

where $\mathbf{\Sigma}_\varepsilon$, a constant matrix, is positive definite.

(b) $\mathbf{x}_{ti}$ is $\mathcal{F}_{t-1}$-measurable and that there exists a non-random positive finite symmetric matrix $\mathbf{B}_T$ for which

$$\mathbf{B}_T^{-1}(\mathbf{Z}\mathbf{Z}')^{1/2} \xrightarrow{p} \mathbf{I}_{mp} \tag{b1}$$

and

$$\max_{1\le t\le T} \|\mathbf{B}_T^{-1}\mathbf{z}_t\| \xrightarrow{p} 0 \tag{b2}$$

as $T \to \infty$. Then the least squares estimate $\widehat{\beta}$ of β has an asymptotic normal distribution in the sense that

$$((\mathbf{Z}\mathbf{Z}')^{1/2} \otimes \mathbf{I}_m)(\widehat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{mp} \otimes \mathbf{\Sigma}_\varepsilon). \tag{A.2}$$

Proof. Some of the arguments in this proof have been adopted from Lai and Wei (1982) and Lai and Robbins (1981). Define

$$\mathbf{x}_{Tt} = \mathbf{B}_T^{-1}\mathbf{z}_t$$

and

$$\widetilde{\mathbf{x}}_{Tt} = \mathbf{x}_{Tt}\mathbb{I}(\|\mathbf{x}_{Tt}\| \le 1) \tag{A.3}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. It follows from (b2) and (A.3) that

$$P(\mathbf{x}_{Tt} \ne \widetilde{\mathbf{x}}_{Tt} \text{ for some } t = 1, ..., T) = P(\|\mathbf{x}_{Tt}\| > 1) \to 0 \tag{A.4}$$

as $T \to \infty$. Furthermore, by Assumption (a) and (A.3), both $\mathsf{E}(\|\widetilde{\mathbf{x}}_{Tt} \otimes \varepsilon_t\|) < \infty$ and $\mathsf{E}(\widetilde{\mathbf{x}}_{Tt} \otimes \varepsilon_t|\mathcal{F}_{t-1}) = \mathbf{0}$ a.s. It follows from (a2), (b1) and (A.4) that

$$\begin{aligned}
&\sum_{t=1}^{T} \mathsf{E}[(\widetilde{\mathbf{x}}_{Tt} \otimes \varepsilon_t)(\widetilde{\mathbf{x}}_{Tt} \otimes \varepsilon_t)'|\mathcal{F}_{t-1}] \\
&= \sum_{t=1}^{T} \mathsf{E}[((\mathbf{B}_T^{-1}\mathbf{z}_t\mathbf{z}_t'\mathbf{B}_T^{-1}) \otimes (\varepsilon_t\varepsilon_t'))\mathbb{I}(\|\mathbf{x}_{Tt}\| \le 1)|\mathcal{F}_{t-1}] \\
&\xrightarrow{p} \mathbf{I}_{mp} \otimes \mathbf{\Sigma}_\varepsilon.
\end{aligned} \tag{A.5}$$

Since $\mathbf{\Sigma}_\varepsilon$ is positive definite, there exists a nonsingular matrix $\mathbf{L}$ such that $\mathbf{L}\mathbf{\Sigma}_\varepsilon\mathbf{L}' = \mathbf{I}_m$. It follows from (a2) that

$$\lim_{t\to\infty} \mathsf{E}\{\mathbf{L}\varepsilon_t\varepsilon_t'\mathbf{L}'|\mathcal{F}_{t-1}\} = \mathbf{I}_m \text{ a.s.} \tag{A.6}$$

Let $\mathbf{d}$ be an arbitrary non-random $(m^2p \times 1)$ vector and any element d_s $(s = 1, ..., m^2p)$ of $\mathbf{d}$ be positive such that $\mathbf{d}'\mathbf{d} = 1$. Then it follows from (A.5)-(A.6) that

$$\sum_{t=1}^{T} \mathsf{E}[\mathbf{d}'(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}|\mathcal{F}_{t-1})] \xrightarrow{p} 1. \tag{A.7}$$

We now have to show that, for any $\delta > 0$, the conditional Lindeberg condition

$$\sum_{t=1}^{T} \mathsf{E}[\mathbf{d}'(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\mathbb{I}\left(\left\|(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\right\|^2 > \delta\right) | \mathcal{F}_{t-1})] \xrightarrow{p} 0 \tag{A.8}$$

holds. Brown (1971, Lemma 2) showed that under condition (A.7), the Lindeberg condition,

$$\sum_{t=1}^{T} \mathsf{E}[\mathbf{d}'(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\mathbb{I}\left(\left\|(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\right\|^2 > \delta\right)] \xrightarrow{p} 0 \tag{A.9}$$

for any $\delta > 0$, and the conditional Lindeberg condition (A.8), are equivalent. Thus we only have to show that (A.9) holds.

For any $\lambda > 0$, it follows from Hall and Heyde (1980, pp. 53-54) that

$$\begin{aligned}
&\sum_{t=1}^{T} \mathsf{E}[\mathbf{d}'(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\mathbb{I}\left(\left\|(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\right\|^2 > \delta\right)] \\
&\leq \mathsf{E}[\sum_{t=1}^{T} \mathbf{d}'(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\mathbb{I}\left(\left(\sum_{t=1}^{T} \left\|(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\right\|^2\right) > \lambda\right)] \\
&+ \lambda P\left(\max_{1 \leq t \leq T}\left(\left\|(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\right\|^2\right) > \delta\right).
\end{aligned} \tag{A.10}$$

Given any $\varepsilon > 0$, choose λ so large that for any T, the first term on the right-hand side of (A.10) is bounded by $\varepsilon/2$. That this can be done follows from (A.7). Now choose T^* sufficiently large such that for all $T \geq T^*$, the second term on the right-hand side of (A.10) is bounded by $\varepsilon/2$. In other words, we have to show that for any $\varepsilon^* > 0$,

$$P\left(\max_{1 \leq t \leq T}\left(\left\|(\widetilde{\mathbf{x}}_{Tt} \otimes \mathbf{L}\varepsilon_t)'\mathbf{d}\right\|^2\right) > \delta\right) < \varepsilon^*. \tag{A.11}$$

Define the i-th element of $\widetilde{\mathbf{x}}_{Tt}$ as $\widetilde{x}_{i,Tt}$ and the j-th element of $\widetilde{\varepsilon}_t = \mathbf{L}\varepsilon_t$ as $\widetilde{\varepsilon}_{jt}$, where $i = 1, ..., mp$ and $j = 1, ..., m$. To show that (A.11) holds, observe that,

for any t,

$$\begin{aligned}
&\mathsf{E}[\left(\sum_{i,j}\widetilde{x}_{i,Tt}\widetilde{\varepsilon}_{jt}d_s\right)^2] \\
&= \mathsf{E}[\mathsf{E}\left(\left(\sum_{i,j}\widetilde{x}_{i,Tt}\widetilde{\varepsilon}_{jt}d_s\right)^2|\mathcal{F}_{t-1}\right)] \\
&\leq \mathsf{E}[\mathsf{E}\left(\left(\sum_{i,j}\widetilde{x}_{i,Tt}\widetilde{\varepsilon}_{jt}\right)^2|\mathcal{F}_{t-1}\right)] \\
&\leq \mathsf{E}[\left(\sum_{i,j}\widetilde{x}_{i,Tt}^2\left(\mathsf{E}\left(\widetilde{\varepsilon}_{jt}^2|\mathcal{F}_{t-1}\right)\right)\right) \\
&+ 2\left(\sum_{i_1\neq i_2 \text{ or } j_1\neq j_2}|\widetilde{x}_{i_1,Tt}\widetilde{x}_{i_2,Tt}|\left(\mathsf{E}\left(|\widetilde{\varepsilon}_{j_1t}\widetilde{\varepsilon}_{j_2t}|\,|\mathcal{F}_{t-1}\right)\right)\right)] \qquad \text{(A.12)}
\end{aligned}$$

where $i_1, i_2 \in \{1, ..., mp\}$ and $j_1, j_2 \in \{1, ..., m\}$. Note that terms on the right-hand of (A.12) satisfy, for any t,

$$\sum_i \widetilde{x}_{i,Tt}^2 \xrightarrow{p} 0$$

and

$$\begin{aligned}
&\sum_{i_1\neq i_2}|\widetilde{x}_{i_1,Tt}\widetilde{x}_{i_2,Tt}| \\
&\leq \sum_{i_1\neq i_2}\left(\max\{|\widetilde{x}_{i_1,Tt}|, |\widetilde{x}_{i_2,Tt}|\}\right)^2 \\
&\xrightarrow{p} 0,
\end{aligned}$$

because (b2) holds. On the other hand, it follows from (a) that for any t and j, $\mathsf{E}\left(\widetilde{\varepsilon}_{jt}^2|\mathcal{F}_{t-1}\right)$ and $\mathsf{E}\left(|\widetilde{\varepsilon}_{j_1t}\widetilde{\varepsilon}_{j_2t}|\,|\mathcal{F}_{t-1}\right)$ are bounded. Then (A.11) holds. Therefore, for all $T \geq T^*$,

$$\sum_{t=1}^{T}\mathsf{E}[\mathbf{d}'(\widetilde{\mathbf{x}}_{Tt}\otimes\mathbf{L}\varepsilon_t)(\widetilde{\mathbf{x}}_{Tt}\otimes\mathbf{L}\varepsilon_t)'\mathbf{d}\mathbb{I}\left(\left\|(\widetilde{\mathbf{x}}_{Tt}\otimes\mathbf{L}\varepsilon_t)'\mathbf{d}\right\|^2 > \delta\right)] < \varepsilon$$

which proves (A.9). Hence by (A.7) and (A.8) and Theorem 2.2 of Dvoretzky (1972),

$$\sum_{t=1}^{T}\mathbf{d}'(\mathbf{B}_T^{-1}\mathbf{z}_t\otimes\mathbf{L}\varepsilon_t) \xrightarrow{d} N(0,1).$$

It follows from the Cram,r-Wold theorem (see, e.g., Davidson, 1994, pp. 206) that

$$\sum_{t=1}^{T}(\mathbf{B}_T^{-1}\mathbf{z}_t \otimes \mathbf{L}\varepsilon_t) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{m^2p}). \tag{A.13}$$

Result (A.13) then yields (A.2). ■

Proof of Theorem. We shall apply the above Lemma to (1.4). Therefore, we have to show that there exists a nonrandom positive definite symmetric matrix $\mathbf{B}_T$ such that $\mathbf{B}_T^{-1}(\mathbf{ZZ}')^{1/2} \xrightarrow{p} \mathbf{I}_{(k+1)(mp+1)}$ and $\max_{1\le t\le T}\left\|\mathbf{B}_T^{-1}\mathbf{z}_t\right\| \xrightarrow{p} 0$.

Let $\mathbf{v}_t = (y_{1,t-1}, \cdots, y_{1,t-p}, \cdots, y_{m,t-1}, \cdots, y_{m,t-p})'$ be an $(mp\times 1)$ random vector such that $\mathbf{w}_t = (1, \mathbf{v}_t')'$ and $\mathsf{E}\mathbf{w}_t = \mu$ and $\mathsf{E}\mathbf{w}_t\mathbf{w}_t' = \mathbf{\Omega}$, where $\mathbf{\Omega}$ is positive definite.

Under H_0 assume that conditions (A1)-(A2) hold. Under those conditions $\{\mathbf{y}_t\}$ in (1.4) is ergodic for the first and second moments. Then, for $i, j = 0, 1, ..., k$,

$$T^{-(i+j+1)}\sum_{t=1}^{T} t^{i+j}\mathbf{w}_t \xrightarrow{p} (i+j+1)^{-1}\mu \tag{A.14}$$

and

$$T^{-(i+j+1)}\sum_{t=1}^{T} t^{i+j}\mathbf{w}_t\mathbf{w}_t' \xrightarrow{p} (i+j+1)^{-1}\mathbf{\Omega} \tag{A.15}$$

as $T \to \infty$. Let $\mathbf{U} = [(i+j+1)^{-1}]$, $i, j = 0, 1, ..., k$, be a $(k+1)\times(k+1)$ matrix and set

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{U} & \mathbf{U}\otimes\mu' \\ \mathbf{U}\otimes\mu & \mathbf{U}\otimes\mathbf{\Omega} \end{bmatrix}.$$

It follows from (A.14) and (A.15) that

$$\operatorname{plim}_{T\to\infty}\mathbf{\Upsilon}_T^{-1}(\mathbf{ZZ}')\mathbf{\Upsilon}_T^{-1} = \mathbf{\Lambda}.$$

Matrix $\mathbf{\Lambda}$ is positive definite because $\mathbf{ZZ}'$ is positive definite. Let $\mathbf{B}_T = (\mathbf{\Upsilon}_T\mathbf{\Lambda}\mathbf{\Upsilon}_T)^{1/2}$. Then

$$\mathbf{B}_T^{-1}(\mathbf{ZZ}')^{1/2} \xrightarrow{p} \mathbf{I}_{(k+1)(mp+1)} \tag{A.16}$$

as $T \to \infty$. Furthermore,

$$\left\|\mathbf{B}_T^{-1}\mathbf{z}_t\right\|^2 = \mathbf{z}_t'\mathbf{\Upsilon}_T^{-1}\mathbf{\Lambda}^{-1}\mathbf{\Upsilon}_T^{-1}\mathbf{z}_t = O_p(T^{-1}), \tag{A.17}$$

because

$$\mathbf{\Upsilon}_T^{-1}\mathbf{z}_t = (\mathbf{D}_k^{-1}\otimes\mathbf{I}_{mp+1})\mathbf{z}_t = T^{-1/2}(\widetilde{\mathbf{s}}_t\otimes\mathbf{w}_t)$$

where $\widetilde{\mathbf{s}}_t = (1, t/T, \cdots, (t/T)^k)'$. Under H_0 it follows from (A.16) and (A.17) that

$$\max_{1 \leq t \leq T} \left\| \mathbf{B}_T^{-1} \mathbf{z}_t \right\| \xrightarrow{p} 0.$$

The error process $\{\varepsilon_t\}$ satisfies the assumptions of Lemma. Application of that lemma yields the result (1.8). ■

Appendix B

Models used in the simulation study

The data-generating TV-VAR model is

$$\mathbf{y}_t = \mu_0 + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{y}_{t-i} + \mathbf{G}(t)(\mu_1 + \sum_{i=1}^{p} \mathbf{\Theta}_i \mathbf{y}_{t-i}) + \varepsilon_t, \; t = 1, \cdots, T \qquad \text{(B.1)}$$

To investigate the size and power properties of the our test four different DGPs were used. The first DGP is a bivariate system (B.1) with $p = 1$, where

$$\begin{aligned} &\mu_0' = \begin{pmatrix} 0.2 & 0.2 \end{pmatrix}, \mu_1' = \begin{pmatrix} 0.23 & 0.23 \end{pmatrix} \\ &\mathbf{\Phi}_1 = \begin{pmatrix} 0.55 & 0.65 \\ -0.038 & 0.3 \end{pmatrix}, \mathbf{\Theta}_1 = \begin{pmatrix} 0.2 & 0.08 \\ -0.03 & -0.2 \end{pmatrix} \end{aligned} \qquad \text{(DGP 1)}$$

The second DGP from van Dijk (1999) is an TV-VAR(2) with $m = 2$ where

$$\begin{aligned} &\mu_0' = \begin{pmatrix} 0.2 & 0.1 \end{pmatrix}, \mu_1' = \begin{pmatrix} 0.22 & 0.11 \end{pmatrix}, \\ &\mathbf{\Phi}_1 = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}, \mathbf{\Phi}_2 = \begin{pmatrix} -0.4 & -0.2 \\ -0.2 & 0.4 \end{pmatrix} \\ &\mathbf{\Theta}_1 = \begin{pmatrix} -0.4 & 0.2 \\ 0.2 & -0.4 \end{pmatrix}, \mathbf{\Theta}_2 = \begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix} \end{aligned} \qquad \text{(DGP 2)}$$

Finally, the last two DGPs considered are a TV-VAR(1) and TV-VAR(2) models with $m = 4$:

$$
\begin{aligned}
\mu_0' &= \begin{pmatrix} 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix} \\
\mu_1' &= \begin{pmatrix} 0.202 & 0.202 & 0.202 & 0.202 \end{pmatrix} \\
\mathbf{\Phi}_1 &= \begin{pmatrix} 0.03 & -0.01 & 0.01 & -0.01 \\ 0.03 & 0.00 & 0.00 & -0.01 \\ 0.00 & 0.01 & 0.03 & -0.01 \\ -0.01 & 0.00 & 0.03 & 0.00 \end{pmatrix} \\
\mathbf{\Theta}_1 &= \begin{pmatrix} 0.80 & 0.00 & 0.05 & 0.00 \\ 0.70 & 0.30 & 0.20 & 0.00 \\ 0.20 & 0.00 & 0.70 & 0.00 \\ 0.00 & 0.30 & 0.70 & 0.30 \end{pmatrix}
\end{aligned}
\qquad \text{(DGP 3)}
$$

and

$$
\begin{aligned}
\mu_0' &= \begin{pmatrix} 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix} \\
\mu_1' &= \begin{pmatrix} 0.21 & 0.21 & 0.21 & 0.21 \end{pmatrix} \\
\mathbf{\Phi}_1 &= \begin{pmatrix} 0.80 & 0.00 & 0.05 & 0.00 \\ 0.70 & 0.30 & 0.20 & 0.00 \\ 0.20 & 0.00 & 0.70 & 0.00 \\ 0.00 & 0.30 & 0.70 & 0.30 \end{pmatrix} \\
\mathbf{\Phi}_2 &= \begin{pmatrix} 0.330 & -0.015 & -0.012 & -0.015 \\ 0.270 & 0.003 & -0.003 & -0.015 \\ -0.003 & -0.015 & 0.027 & -0.015 \\ -0.015 & 0.003 & 0.027 & 0.003 \end{pmatrix} \\
\Theta_1 &= \begin{pmatrix} -0.9680 & 0.000 & -0.0605 & 0.000 \\ -0.847 & -0.3630 & -0.2420 & 0.000 \\ -0.2420 & 0.000 & -0,8470 & 0.000 \\ 0.000 & -0.3630 & -0.8470 & -0,3630 \end{pmatrix} \\
\Theta_2 &= \begin{pmatrix} -0.0330 & 0.0150 & 0.012 & 0.015 \\ -0,0270 & -0.003 & 0.003 & 0.015 \\ 0.0030 & 0.0150 & -0,027 & 0.015 \\ 0.0150 & -0.0030 & -0.027 & -0.003 \end{pmatrix}
\end{aligned}
\qquad \text{(DGP 4)}
$$

To assess the power properties of the tests, we assume that the transition function (1.2) is equal for all the equations, with $\gamma = 0, 10$ or 20, $c = (0.5)\,T$ for $k = 1$ and $\mathbf{c} = ((0.3)\,T, (0.7)\,T)$ for $k = 2$. Additionally, the sample size was either $T = 50$ or $T = 100$.

In power comparisons with the generalized Chow-test the model is a bivariate system with $p = 1$. The model under the alternative hypothesis is (B.1) with the following parameter values for $\gamma > 0$:

$$\mu_0' = \begin{pmatrix} 0.2 & 0.2 \end{pmatrix}, \mu_1' = \begin{pmatrix} 0.23 & 0.23 \end{pmatrix}$$
$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.7 & 0.2 \\ -0.2 & 0.7 \end{pmatrix}, \mathbf{\Theta}_1 = \begin{pmatrix} -0.4 & 0.2 \\ 0.2 & -0.4 \end{pmatrix}$$

When $\gamma = 0$, (B.1) is linear with parameters

$$(\mu_0^*)' = \mu_0' + 0.5\mu_1' = \begin{pmatrix} 0.315 & 0.315 \end{pmatrix}$$
$$\mathbf{\Phi}_1^* = \mathbf{\Phi}_1 + 0.5\mathbf{\Theta}_1 = \begin{pmatrix} 0.5 & 0.3 \\ -0.1 & 0.5 \end{pmatrix}. \qquad \text{(DGP 2b)}$$

This parameterization is used for obtaining the size results in Table 3.

Finally, to investigate the power of the tests, two different alternative models are used. When the model under the alternative is DGP 1b, we assume that the transition function (1.2) is the same for all the equations, with $k = 1$, $\gamma = 5$, 10 and $c = 0.5$. This model is used for obtaining the power results in Table 3. On the other hand, when the alternative is a structural break, we consider two additional parameter generated form DGP 2b as follows:

$$\mu_1 = \begin{pmatrix} 0.2 \\ 0.2 \end{pmatrix} - \alpha \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}$$
$$\mathbf{\Theta}_1 = \begin{pmatrix} 0.5 & 0.3 \\ 0.4 & 0.5 \end{pmatrix} - \alpha \begin{pmatrix} 1.0 & 1.0 \\ 1.0 & 1.0 \end{pmatrix}$$

where $\alpha = 0.15, 0.30$.

In all cases, ε_t are independent normally distributed with mean zero and covariance matrix $\mathbf{\Sigma} = [\sigma_{ij}]$ with $\sigma_{ii} = 0.3$ and $\sigma_{ij} = 0.2$ for $i \neq j$.

Chapter 2

Panel smooth transition regression model and an application to investment under credit constraints

2.1 Introduction

A standard assumption in panel models is that regression functions are identical across all observations in a sample. In some applications this assumption may be violated. Hansen (1999a) suggested an alternative in which the individual observations are divided into classes according to an observable variable. If the regression functions are the same for all observations in a class and this holds for all classes, this assumption leads to a panel threshold regression (PTR) model. Hansen (1999a) derives econometric techniques for the PTR model. These include maximum likelihood estimation, asymptotic confidence intervals for the parameters as well as specification tests for determining the number of classes. The last problem leads to nonstandard inference because of an identification problem present in the tests. The PTR approach is motivated by an empirical example, in which there exists theory suggesting that the regression functions may not be identical across samples. In particular, the economic theory suggests that in the case of imperfect information, external finance may be limited, and already heavily indebted firms may have to use their cash flow to finance their investments. This separates them from other firms whose access to external sources of financing is not restricted. A leading article making use of this classification of firms is Fazzari, Hubbard, and Petersen (1988).

At the end of his article, Hansen (1999a) points out that one could also apply models with a smooth transition to this problem. In that case, instead of a small finite number of classes there would be a smooth transition controlled by an observable variable from one extreme regime to another. In the application to investment financing, that would mean that the degree of indebtedness would have a more subtle effect on the availability of external financing than a PTR model would allow. In this chapter we introduce a nondynamic fixed effect panel model in which the regression coefficients are allowed to change smoothly as a function of an exogenous variable. In this sense, the chapter offers an alternative methodology to Hansen (1999a).

This chapter is organized as follows: The next section introduces the panel smooth transition model. The third section discusses estimation and model building. Section 4 contains small-sample properties of the specification procedure. Section 5 contains an illustration of the proposed methodology where it is applied to the economic problem analyzed by Hansen (1999a). Section 6 concludes.

2.2 Panel smooth transition regression model

The Panel Smooth Transition Regression (PSTR) model is a fixed effect model with exogenous regressors. The basic PSTR model with a single transition is defined as follows:

$$y_{it} = \mu_i + \beta_1' x_{it} + \beta_2' x_{it} g\left(q_{it}; \gamma, \mathbf{c}\right) + u_{it} \tag{2.1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The dependent variable, y_{it}, is a scalar, μ_i is an unobservable time-invariant regressor, x_{it} is a $k-$dimensional vector of time-varying exogenous variables[1], q_{it} is an observable transition variable and u_{it} are the errors.

The main feature of this model is the transition function $g\left(q_{it}; \gamma, \mathbf{c}\right)$. It is a continuous and bounded function of q_{it} that allows the parameter in (2.1) to change smoothly as a function of q_{it}. In this work we follow Granger and Teräsvirta (1993), Teräsvirta (1994) and Jansen and Teräsvirta (1996) and define

$$g\left(q_{it}; \gamma, \mathbf{c}\right) = \left(1 + \exp\left(-\gamma \prod_{j=1}^{m} (q_{it} - c_j)\right)\right)^{-1} , \gamma > 0, c_1 \leq, \ldots, \leq, c_m \tag{2.2}$$

where $\mathbf{c} = (c_1, \ldots, c_m)'$ is an m-dimensional vector of location parameters, and $\gamma \geq 0$ and $c_1 \leq \cdots \leq c_m$ are identification restrictions. Parameter γ determines the slope of the transition function. When $m = 1$ and $\gamma \to \infty$, (2.1) and (2.2) define the two-regime PTR model in Hansen (1999a). When $m > 1$ and $\gamma \to \infty$, the number of identical regimes remains two, but the function switches between zero and one at $c_1, \ldots, c_m$. Finally, when $\gamma \to 0$, the transition function (2.2) becomes constant and the model is the standard linear model with fixed effects.

The transition function (2.2) with $m = 1$ or $m = 2$ is already a very flexible parametrization since it allows different types of changes in the parameters. For example, if $m = 2$, $c_1 = c_2 = c$, (2.2) implies that only the Euclidean distance between q_{it} and c has an effect on y_{it}. Moreover, if $\gamma \to \infty$, transition function (2.2) defines a three-regime model whose outer regimes are identical and different from the mid-regime. Finally, when $m = 1$, the model allows a single monotonic smooth transition whose location is controlled by c_1.

[1]Lagged values of the dependent variable are not allowed because the presence of the fixed effect would invalidate the use of the within transformation to handle the nuisance parameters; see Chamberlain (1984, p 1256).

A possible generalization of the PSTR model is the general additive PSTR model

$$y_{it} = \mu_i + \beta_0' x_{it} + \sum_{j=1}^{r} \beta_j' x_{it} g_j(q_{it}^{(j)}; \gamma_j, \mathbf{c}_j) + u_{it} \tag{2.3}$$

where the transition functions are of type (2.2). If $m = 1$ for g_j, $j = 1, \ldots, r$, $q_{it}^{(j)} \equiv q_{it}$ and $\gamma_j \to \infty$, $j = 1, \ldots, r$, (2.3) collapses into an $(r+1)$-regime PTR model of Hansen (1999a). Consequently, the general additive PSTR model can be used as an alternative to multiple-regime PTR model. Additionally, when the larger model the investigator is willing to consider is a PSTR model (2.1) with $r = 1$ and $m = 1$ or $m = 2$, model (2.3) can serve as an alternative in the evaluation of the estimated PSTR model (2.1). This possibility will be discussed in Section 2.3.4

2.3 Building panel smooth transition regression models

2.3.1 Modelling cycle

The PSTR model is a nonlinear model, and its use requires a systematic modelling strategy. Hansen (1999a) outlines a modelling cycle for the PTR model that consists of testing linearity and selecting the number of regimes using statistical tests. The latter stage also implies maximum likelihood estimation of the parameters of at least two PTR models. Hansen assumes that the threshold variable is given, but if it were not, his procedure could probably be extended to include the possibility of choosing it from a set of candidate variables as in Hansen (1999b).

In this chapter, we consider a modelling cycle for PSTR models consisting of specification, estimation and evaluation stages. Specification includes testing linearity and, if it is rejected, determining the form of the transition function (2.2), that is, choosing between $m = 1$ and $m = 2$. At the evaluation stage the estimated model is subjected to misspecification tests to check whether or not it can be considered an adequate description of the data. The null hypotheses to be tested include parameter constancy, no remaining nonlinearity and no autocorrelation in the errors.

A similar cycle has been previously suggested for smooth transition autoregressive (STAR) and smooth transition regression (STR) models; see, for example, Teräsvirta (1998) or van Dijk, Teräsvirta, and Franses (2002) for description and discussion. It has inspired the techniques developed in the present work.

2.3.2 Testing linearity against PSTR model

The first step of the specification stage is to test linearity against PSTR. This is important for both statistical and economic reasons. Statistically, the PSTR model is not identified if the data-generating process is linear, and a linearity test is necessary to avoid the estimation of unidentified models. The PTR model has the same property. From the economics point of view, a linearity test may account for testing some economic theory suggestions. For instance, in the example on the access of firms to external financing, established theory suggests a linear model, whereas a nonlinear model is required if there are credit restrictions that depend on the degree of indebtedness of the firm.

Testing linearity in the PSTR model (2.1) can be carried out in two ways either by testing $\mathrm{H}_0^1 : \beta_2 = 0$ or $\mathrm{H}_0^2 : \gamma = 0$. In both cases the test will be nonstandard because under either null hypothesis, the PSTR model contains unidentified nuisance parameters. In particular, $(\gamma, \mathbf{c}')$ are not identified under H_0^1 and $(\beta_2', \mathbf{c}')$ under H_0^2. The testing problem when unidentified nuisance parameters are presented under the null was first studied by Davies (1977, 1987). Luukkonen, Saikkonen, and Teräsvirta (1988), Andrews and Ploberger (1994) and Hansen (1996) proposed alternative solutions to the problem. Recently, Hansen (1999a, 2000) applied his testing procedure in the PTR framework. We follow Luukkonen, Saikkonen, and Teräsvirta (1988) and test the linearity hypothesis as $\mathrm{H}_0 : \gamma = 0$. To circumvent the identification problem, we replace $g(q_{it}; \gamma, \mathbf{c})$ by its first-order Taylor expansion around $\gamma = 0$ and test an equivalent hypothesis in an auxiliary regression. After replacing $g(q_{it}; \gamma, \mathbf{c})$ in (2.1) by its Taylor expansion and merging terms we obtain the following auxiliary regression,

$$y_{it} = \mu_i + \beta_1'^* x_{it} + \beta_2'^* x_{it} q_{it} + \cdots + \beta_m'^* x_{it} q_{it}^m + u_{it}^* \tag{2.4}$$

where the parameter vectors $\beta_2^*, \ldots, \beta_m^*$ are multiples of γ and $r_{it} = u_{it} + O(\gamma^m)\beta_2' x_{it}$. Testing $\mathrm{H}_0 : \gamma = 0$ in (2.1) is equivalent to testing $\mathrm{H}_0^* : \beta_2^* = \cdots = \beta_m^* = 0$ in (2.4). Note that under the null hypothesis $\{u_{it}^*\} = \{u_{it}\}$, so the Taylor series approximation does not affect the asymptotic distribution theory.

We make the following assumptions about model (2.1) under the null hypothesis:

Assumption L1: For each t, $\{y_{it}, x_{it}, q_{it}\}$ are independently distributed (i.d.) across i.

Assumption L2: For each i, u_{it} is i.i.d over t and independent of $\{(x_{it}, q_{it})_{t=1}^T\}$,

and $\mathsf{E}(u_{it}) = 0$.

Assumption L3: $\mathsf{E}|x_{its}q_{it}^{m}|^{1+\delta} \leq \Delta_1 < \infty$, for $i = 1,\ldots,N$, $t = 1,\ldots,T$, $s = 1,\ldots,k$, where $\delta > 1$.

Assumption L4: $\mathsf{E}|x_{its}q_{it}^{m}u_{it}|^{2+\delta} \leq \Delta_2 < \infty$ for $i = 1,\ldots,N$, $t = 1,\ldots,T$, $s = 1,\ldots,k$, where $\delta > 1$.

Theorem 1 *If assumptions L1 to L4 are satisfied, then the least squared estimator $\hat{\beta}$ of $\beta = (\beta_1^*,\ldots,\beta_m^*)'$ is consistent and asymptotically normal under the null hypothesis when $N \to \infty$ with T fixed.*

Proof. See Appendix A.1 ■

Even though the null hypothesis H_0^* can be tested using any of the three classical tests, we restrict ourselves to the LM test because it only requires the estimation of (2.4) under the null. The computation of the LM statistic involves two steps. First, eliminate the fixed effect from (2.4). Second, compute the LM statistic for the transformed model. The LM test and its F-version can be computed in three stages as follows:

1. Regress $\tilde{y}_{it} = y_{it} - \sum_t y_{it}/T$ on $\tilde{x}_{it} = x_{it} - \sum_t x_{it}/T$ and compute the sum of squared residuals SSR_0.

2. Regress $\tilde{y}_{it}$ on $\tilde{x}_{it}$ and $(x_{it}'q_{it} - \sum_t x_{it}'q_{it}/T,\ldots,x_{it}'q_{it}^{m} - \sum_t x_{it}'q_{it}^{m}/T)$ and compute the sum of squared residuals SSR_1.

3. Compute,

$$\mathrm{LM} = TN(\mathrm{SSR}_0 - \mathrm{SSR}_1)/\mathrm{SSR}_0 \tag{2.5}$$

$$\mathrm{LM_F} = \{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/mk\} \,/\, \{\mathrm{SSR}_1/(TN - N - mk)\} \tag{2.6}$$

Under the null hypothesis, statistic (2.5) is asymptotically distributed as χ_{mk}^2 and the F-statistic (2.6) has an approximate $\mathrm{F}\,[mk, TN - N - mk]$ distribution.

Suppose that the larger model the investigator is willing to consider is the PSTR model (2.1) with $m = 1$ or 2 in (2.2). The linearity test can then be used to choose between $m = 1$ and $m = 2$. Granger and Teräsvirta (1993) and Teräsvirta (1994) proposed the use of a sequence of linearity tests for determining m. The testing sequence applied to the present situation is the following: Using the auxiliary regression (2.4) with $m = 3$, test the null hypothesis $\mathrm{H}_0 : \beta_3 = \beta_2 = \beta_1 = 0$. If it is rejected, test $\mathrm{H}_{04} : \beta_3 = 0$, $\mathrm{H}_{03} : \beta_2 = 0|\beta_3 = 0$ and $\mathrm{H}_{02} : \beta_1 = 0|\beta_3 = \beta_2 = 0$. Select $m = 2$ if the rejection of H_{03} is the strongest one, and otherwise select $m = 1$. For the reasoning behind this rule, see Teräsvirta (1994).

2.3.3 Estimation of parameters

Estimation of parameters of the PSTR model (2.1) is a relatively straightforward application of the fixed effect estimator and nonlinear least squares [NLS]. One has to eliminate the individual effects μ_i by removing individual-specific means and to apply NLS to the transformed model to estimate the remaining parameters. This estimating procedure can be seen as maximum likelihood where first the likelihood function is concentrated with respect to the fixed effects.

In order to discuss the asymptotic properties of the ML estimator we write (2.1) for individual i as,

$$Y_i = \iota\mu_i + X_{1i}\beta_1 + W_i(\beta_3)\beta_2 + U_i \tag{2.1'}$$

where ι is a $(T \times 1)$ vector of ones, $X_{1i} = (x'_{i1}, \ldots, x'_{iT})'$ and $W_i(\beta_3) = (g(X_{2i}\beta_3) \odot X_{1i})$ with $X_{2i} = (\iota', q'_i, \ldots, q_i^{'m})'$ and $\beta_3 = \gamma(1, c_1^*, \ldots, c_m^*)'$. The dimensions of X_{1i}, X_{2i} are $(T \times k)$, $(T \times (m+1))$, respectively.

We make the following assumptions for the PSTR model (2.1) or (2.1'):

Assumption E1: $\{Y_i, X_{1i}, q_i\}$ is an independently identically distributed sequence of random variables and $U_i = Y_i - \mathsf{E}[Y_i|\mu_i, X_{1i}, q_i]$.

Assumption E2: $g(x'_{2it}\beta_3)\beta_2 - g(x'_{2it}\beta_3^0)\beta_2^0 \neq 0$ when $\beta_2 \neq \beta_2^0$ and/or $\beta_3 \neq \beta_3^0$.

Assumption E3: The parameter space Θ is a compact subset of $\mathbb{R}^K$ and $\beta^0 \in \Theta$.

Assumption E4: $\mathsf{E}[|u_{it}^2|^2] \leq \Delta_1 < \infty$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$.

Assumption E5: $\mathsf{E}[|x_{2it,j}x_{2is,h}x_{is,r}x_{it,l}|^2] \leq \Delta < \infty$, for $j, h = 1, \ldots, m$, $r, l = 1, \ldots, k$ and $i = 1, \ldots, N$, $t = 1, \ldots, T$.

Assumption E6: $\mathsf{E}[|x_{jit,h}|^2] \leq \Delta < \infty$, for $j = 1, 2$, $i = 1, \ldots, N$, $h = 1, \ldots, k$ and $t = 1, \ldots, T$.

Assumption E7: $V \equiv \mathsf{E}\left[[X_{1i} \vdots W_i(\beta_3^0)]' Q_T [X_{1i} \vdots W_i(\beta_3^0)]\right]$ is positive definite. $Q_T = I_T - \frac{1}{T}\iota\iota'$ is the within transformation matrix.

Theorem 2 *If assumptions (E1) to (E7) are satisfied, the maximum likelihood estimator is consistent and asymptotically normal when $N \to \infty$ and T is fixed.*

Proof. See Appendix A.2 ■

The only assumption that is not standard is (E2) which is an identification assumption, assumptions (E1), (E3) to (E7) are standard in linear panel models with strictly exogenous regressor. Even though we have assumed that the observations across individuals are i.i.d it could be relaxed in order to allow for heterogeneity. Such a generalization would imply the existence of higher-order moments. [see White (1980) and White (2000) for details].

As mentioned before, the estimation of the parameters in (2.1) is carried out in two steps. First, we eliminate the fixed effects and then apply NLS to the transformed model. Even though the first step is standard in linear models, equation (2.1) calls for a more careful treatment. Specifically, note that the individual means in (2.1) have the form

$$\bar{y}_i = \mu_i + \beta_1 \bar{x}_i + \beta_2 \bar{w}_i (\gamma, \mathbf{c}) + \bar{u}_i \tag{2.7}$$

where $\bar{y}_i$, $\bar{x}_i$, $\bar{w}_i$ and $\bar{u}_i$ are individual means. Subtracting equation (2.7) from equation (2.1) yields

$$\tilde{y}_{it} = \beta' \tilde{x}^*_{it} (\gamma, \mathbf{c}) + \tilde{u}_{it} \tag{2.8}$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{u}_{it} = u_{it} - \bar{u}_i$, $\beta = (\beta_1', \beta_2')'$, $\tilde{x}^*_{it}(\gamma, \mathbf{c}) = (x_{it}' - \bar{x}_i', x_{it}' g(q_{it}; \gamma, \mathbf{c}) - \bar{w}_i'(\gamma, \mathbf{c}))'$. Consequently, the transformed vector $\tilde{x}^*_{it}(\gamma, \mathbf{c})$ in (2.8) depends on $(\gamma, \mathbf{c}')'$ through both the levels and the individual means. For this reason, $\tilde{x}^*_{it}(\gamma, \mathbf{c})$ has to be recomputed at each iteration.

The iterations have the following form. First, given $\left(\gamma^{(j)}, \mathbf{c}^{(j)\prime}\right)'$ estimate $\beta^{(j)}$ by ordinary least squares, which yields

$$\hat{\beta}^{(j)} = \left(\sum_i \sum_t \tilde{x}^*_{it} \left(\gamma^{(j)}, \mathbf{c}^{(j)}\right) \tilde{x}^{*\prime}_{it} \left(\gamma^{(j)}, \mathbf{c}^{(j)}\right) \right)^{-1} \sum_i \sum_t \tilde{x}^*_{it} \left(\gamma^{(j)}, \mathbf{c}^{(j)}\right) y^*_{it} \tag{2.9}$$

Then, conditionally on $\beta^{(j)}$, estimate $\left(\gamma^{(j+1)}, \mathbf{c}'^{(j+1)}\right)'$ by NLS. This amounts to solving

$$\left(\hat{\gamma}^{(j+1)}, \hat{\mathbf{c}}'^{(j+1)}\right)' = \arg\min_{(\gamma, \mathbf{c})} \sum_{i=1} \sum_{t=1} \left(\bar{y}_{it} - \beta'^{(j)} \tilde{x}^*_{it} (\gamma, \mathbf{c})\right)^2 \tag{2.10}$$

Leybourne, Newbold, and Vougas (1998) proposed a similar procedure for STAR models; see also Teräsvirta (1998) and van Dijk, Teräsvirta, and Franses (2002) for discussion. Small-sample properties of this procedure are investigated by simulation in Section 2.4.1.

An issue that deserves special attention in the estimation of the PSTR model is the selection of starting-values. Good starting values may considerably facilitate the numerical optimization or, conversely, inappropriate

starting-values may cause problems. A feasible method for smooth transition models is a grid search. It is seen that (2.8) is linear in parameters when parameters $(\gamma, \mathbf{c}')'$ are fixed. This suggest the following algorithm. First, define an array of values for $(\gamma, \mathbf{c}')'$ such that $\gamma > 0$, and $c_{j\max} < \max\{q_{it}\}$ and $c_{j\min} > \min\{q_{it}\}$, $j = 1, \ldots, m$. Calculate (2.9) for all these values in turn and select the vector $(\gamma^*, \mathbf{c}'^*)'$ minimizing the sum of squared residuals as starting-values of the estimation algorithm. Hansen (1999a) also applied a form of grid search in the estimation of the parameters of PTR model.

2.3.4 Evaluation of the estimated model

After estimating the parameters, the estimated PSTR model has to be evaluated. In this section we consider a number of misspecification tests for this purpose. One of them, the test of no remaining nonlinearity, may also be viewed as a specification test. In this test, the alternative hypothesis is a multiple PSTR model, and the test is thus a smooth transition counterpart to the test in Hansen (1999a) for determining the number of regimes in the PTR model.

The tests to be considered in this section resemble the ones that Eitrheim and Teräsvirta (1996) derived for STAR models. It turns out that they can be modified to fit the present framework. The new tests are the test of parameter constancy and that of no remaining nonlinearity. Error autocorrelation is also an indicator of misspecification. Its presence can, however, already be tested by applying the test by Baltagi and Li (1995).

Testing parameter constancy

Testing parameter constancy in panel data models has not received as much attention as in the time series literature. A possible explanation is that in many applications T is relatively small, which makes the assumption of parameter constancy difficult to test. However, with an increasing number of panels with relatively large T the test for parameter constancy in fixed effects models becomes feasible and important. Even though our test is developed for PSTR models, after minor modifications it can be applied to linear fixed effects models.

Our alternative to parameter constancy is that the parameters in (2.1) change smoothly from one regime to another. The model under the alternative may be called the Time Varying Panel Smooth Transition regression [TV-

PSTR] model. It can be written as follows:

$$y_{it} = \mu_i + \left[\beta_{11}' x_{it} + \beta_{12}' x_{it} g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)\right] + f\left(t/T; \gamma_2, \mathbf{c}_2\right)\left[\beta_{21}' x_{it} + \beta_{22}' x_{it} g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)\right] + u_{it} \tag{2.11}$$

where $g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)$ and $f\left(t/T; \gamma_2, \mathbf{c}_2\right)$ are transition functions as defined in (2.2) and $\left(\gamma_1, \mathbf{c}_1'\right)'$ and $\left(\gamma_2, \mathbf{c}_2'\right)'$ are the parameter vectors. Equation (2.11) has a structure similar to the time-varying STAR model discussed in Lundbergh, Teräsvirta, and van Dijk (2003). One may also write (2.11) as follows:

$$y_{it} = \mu_i + \left[\beta_{11} + \beta_{12} f\left(t/T; \gamma_2, \mathbf{c}_2\right)\right]' x_{it} + \left[\beta_{21} + \beta_{22} f\left(t/T; \gamma_2, \mathbf{c}_2\right)\right]' x_{it} g\left(q_{it}; \gamma_1, \mathbf{c}_1\right) + u_{it}. \tag{2.12}$$

Equation (2.12) shows how the parameters of the model vary between β_{11} and $\beta_{11} + \beta_{12}$ and β_{21} and $\beta_{21} + \beta_{22}$, respectively, smoothly and deterministically over time.

The alternative model (2.11) allows multiple alternatives to parameter constancy depending on the specification of $f\left(t/T; \gamma_2, \mathbf{c}_2\right)$. The general specification of $f\left(t/T; \gamma_2, \mathbf{c}_2\right)$ is

$$f\left(t/T; \gamma_2, \mathbf{c}_2\right) = \left(1 + \exp\left(-\gamma_2 \prod_{j=1}^{h} (t/T - c_{2j})\right)\right)^{-1}, t = 1, \ldots, T \tag{2.13}$$

where $\mathbf{c}_2 = (c_{21}, \ldots, c_{2h})'$ is an h-dimensional vector of location parameters. As before, the value of h determines the alternative hypothesis. In particular, it controls the form of switches in parameters. When $h = 1$, the TV-PSTR model allows monotonic change in parameters. Equivalently, when $h = 2$, the parameters change symmetrically around $(c_{21} + c_{22})/2$. Finally, γ_2 measures the smoothness of the change: when $\gamma_2 \to \infty$ in (2.13), $f\left(t/T; \gamma_2, \mathbf{c}_2\right)$ becomes a step function, so structural breaks are included in the alternative as special cases. On the other hand, when $\gamma_2 = 0$ in (2.13), model (2.11) has constant parameters.

Note that it is assumed that the parameters in the transition function $g(q_{it}; \gamma, c_1)$ are constant over time. This assumption is a practical one: such "second-order" nonconstancy is considerably harder to detect than nonconstancy in the regression coefficients, in particular as T may not be large in applications. We also assume a common transition function $f\left(t/T; \gamma_2, \mathbf{c}_2\right)$ for all individuals.

The null hypothesis of constant parameters in model (2.11) can be stated as $\mathrm{H}_0 : \gamma_2 = 0$. However, under this hypothesis $\left(\beta_{12}', \beta_{22}', \mathbf{c}_2'\right)'$ are not identified.

To circumvent this problem we follow Eitrheim and Teräsvirta (1996) and replace (2.13) in (2.11) by its first-order Taylor expansion around $\gamma_2 = 0$. After merging terms we get the following auxiliary regression

$$\begin{aligned} y_{it} &= \mu_i + x'_{it}\beta^*_{11} + x'_{it}(t/T)\beta^*_1 + \cdots + x'_{it}(t/T)^h\beta^*_h \\ &+ \left\{x'_{it}\beta^*_{21} + x'_{it}(t/T)\beta^*_{h+1} + \cdots + x'_{it}(t/T)^h\beta^*_{2h}\right\} g\left(q_{it};\gamma_1,\mathbf{c}_1\right) + u^*_{it} \end{aligned} \tag{2.14}$$

where $u^*_{it} = u_{it} + R\left(t/T,\gamma_2,\mathbf{c}_2\right)$ and $R\left(t/T,\gamma_2,\mathbf{c}_2\right)$ is the approximation error in the Taylor expansion. In (2.14), $\beta^*_j = \gamma_2\beta_j$ for $j = 1,2,\ldots,h,h+1,h+2,\ldots,2h$. Then the original null hypothesis, $\mathrm{H}_0 : \gamma_2 = 0$ can be tested in the auxiliary regression (2.14) as $\mathrm{H}'_0 : \beta^*_j = 0$ for $j = 1,2,\ldots,h,h+1,h+2,\ldots,2h$. Finally, note that under H'_0 $u^*_{it} = u_{it}$, $i = 1,\ldots,N$, so the Taylor series approximation does not affect the distribution assumptions. Therefore, under the null hypothesis and standard regularity conditions, the NLS estimator of $\mathbf{b} = (\beta^{*\prime}_{11},\beta^{*\prime}_1,\ldots,\beta^{*\prime}_{21},\beta^{*\prime}_{h+1},\ldots,\beta_{2h},\gamma,\mathbf{c}')'$ is consistent and asymptotically normal for fixed T and $N \to \infty$.

In this context it is convenient to use the LM test because it only requires the estimation of (2.14) under the null hypothesis. In order to compute the LM statistic and its F-version we need to define the following vectors:

$$\begin{aligned} \hat{v}_{it} &= \left(\tilde{w}'_{it},\tilde{z}'_{it},\left(\partial\tilde{z}_{it}/\partial\gamma_1\right)'\hat{\beta}_{12},\left(\partial\tilde{z}_{it}/\partial c_{11}\right)'\hat{\beta}_{12},\ldots,\left(\partial\tilde{z}_{it}/\partial c_{1m}\right)'\hat{\beta}_{12}\right)' \\ \hat{\xi}_{it} &= (\tilde{x}'_{it},\tilde{\psi}'_{it},\left(\partial\tilde{\psi}_{it}/\partial\gamma_1\right)',\left(\partial\tilde{\psi}_{it}/\partial c_{11}\right)',\ldots,\left(\partial\tilde{\psi}_{it}/\partial c_{1m}\right)')' \\ \tilde{\psi}_{it} &= x_{it}g\left(q_{it},\gamma_1,\mathbf{c}_1\right) - 1/T\sum_{t=1}^{T} x_{it}g(q_{it},\gamma_1,\mathbf{c}_1) \\ \tilde{w}_{it} &= x_{it}(t/T)^j - 1/T\sum_{t=1}^{T} x_{it}(t/T)^j, j = 1,\ldots,h \end{aligned}$$

and

$$\tilde{z}'_{it} = x_{it}g\left(q_{it};\gamma_1,\mathbf{c}_1\right)(t/T)^j - 1/T\sum_{t=1}^{T} x_{it}g\left(q_{it};\gamma_1,\mathbf{c}_1\right)(t/T)^j, j = 1,\ldots,h$$

The χ^2 and F version of the test can be computed in three stages as follows[2]:

[2]See Appendix B for the mathematical derivation of the test.

1. Estimate the PSTR model and compute the residual sum of squares SSR_0.

2. Regress $\tilde{y}_{it}$ on $\hat{v}_{it}$ and $\hat{\xi}_{it}$ and compute the residual sum of squares SSR_1.

3. Compute the χ^2 and F-versions of the tests as follows;

$$\begin{aligned} \mathrm{LM} &= TN(\mathrm{SSR}_0 - \mathrm{SSR}_1)/SSR_0 \\ \mathrm{LM_F} &= \{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/2hk\} \,/\, \{\mathrm{SSR}_1/(TN - N - 2hk)\} \end{aligned} \tag{2.15}$$

   Under the null hypothesis LM is asymptotically distributed as $\chi^2_{(2hk)}$ and $\mathrm{LM_F}$ is approximately distributed as $F(2hk, TN - N - 2hk)$.

Small-sample properties of $\mathrm{LM_F}$ will be investigated by simulation in Section 2.4.2.

Testing the hypothesis of no remaining nonlinearity

The purpose of this test is twofold. First, if the basic PSTR model (2.1) with (2.2) is the largest model that one wants to consider, the test is a misspecification test. A rejection indicates that the specification is not satisfactory. If the investigator is willing to consider a multiple PSTR model, the test serves as a test for the number of transition functions in the model. When the test is carried out, it is not necessary to assume that the new transition function in the alternative has the same transition variable as the one in the estimated model. As already mentioned, the test bears resemblance to the test of a similar hypothesis in Eitrheim and Teräsvirta (1996).

The model we consider in this subsection is the general additive PSTR model

$$y_{it} = \mu_i + x'_{it}\beta_0 + \sum_{j=1}^{r} g_j(q_{it}^{(j)}; \gamma_j, c_j)x'_{it}\beta_j + u_{it} \tag{2.16}$$

where r is the number of regimes. This model can be written in a way that resembles the Multiple Regime Threshold model presented by Hansen (1999a). In fact, after adding and subtracting the appropriate elements, (2.16) becomes an $(r+1)$-regime STR panel model,

$$y_{it} = \mu_i + \left(1 - \sum_{j=1}^{r} g\left(q_{it}^{(j)}; \gamma_j, c_j\right)\right) x'_{it}\beta_0 + \sum_{j=1}^{r} x'_{it}\left(\beta_0 + \beta_j\right) g\left(q_{it}^{(j)}; \gamma_j, \mathbf{c}_j\right) + u_{it}$$

where $g\left(q_{it}^{(j)}; \gamma_j, \mathbf{c}_j\right)$ are transition functions as defined in (2.2) with $m = 1$.

Consider the case in which $r = 2$ and $q_{it}^{(j)} = q_{it}$ for $j = 1, 2$. Equation (2.16) then becomes three regime STR panel model of the form:

$$y_{it} = \mu_i + (1 - g_1 - g_2)\, x_{it}'\beta_0 + g_1 x_{it}'\,(\beta_0 + \beta_1) + g_2 x_{it}'\,(\beta_0 + \beta_2) + u_{it} \quad (2.17)$$

Two regimes in (2.17) are associated with $g_1 = g_2 = 0$ and $g_1 = g_2 = 1$, respectively, and there is an intermediate regime associated with $g_1 = 0$ and $g_2 = 1$ or $g_1 = 1$ and $g_2 = 0$. When $\gamma_j \to \infty$, $j = 1, 2$, model (2.17) collapses into a three-regime PTR model of Hansen (1999a).

The test of no remaining nonlinearity can also be used as a test for determining the number of transition functions in the PSTR model. In applications a linear panel model is often applied by estimating several models into which the observations are allocated by a classifier q_{it}. In general, the number of models or regimes and the values of q_{it} that define the different models are selected on an *ad hoc* basis. Here we show how one can select the number of regimes sequentially.

In order to demonstrate the procedure, we assume that a PSTR model (2.16) with $r = 1$ has been estimated and adding another transition function is considered. The extended model can be written as follows:

$$y_{it} = \mu_i + x_{it}'\beta_0 + g_1\,(q_{it}; \gamma_1, c_1)\, x_{it}'\beta_1 + g_2\,(q_{it}; \gamma_2, c_2)\, x_{it}'\beta_2 + u_{it} \quad (2.18)$$

and the null hypothesis of no additional transition in (2.18) is $H_0 : \gamma_2 = 0$. Under H_0, the parameters in (2.18) cannot be estimated consistently because the model is not identified. As before, the identification problem can be circumvented by replacing $g_2\,(q_{it}; \gamma_2, c_2)$ in (2.18) by a Taylor expansion around $\gamma_2 = 0$. Choosing a first-order Taylor approximation leads to testing the hypothesis $H_0' : \beta_{22}^* = \cdots = \beta_{2m}^* = 0$ in the following auxiliary regression:

$$\begin{aligned} y_{it} \;=\; & \mu_i + x_{it}'\beta_0^* + g_1\,(q_{it}; \hat{\gamma}_1, \hat{c}_1)\, x_{it}'\beta_1 \\ & + (x_{it}q_{it})'\,\beta_{22}^* + \cdots + (x_{it}q_{it}^m)'\,\beta_{2m}^* + e_{it}^* \end{aligned} \quad (2.19)$$

where $(\gamma_1, c_1')'$ is the parameter vector estimated under the null hypothesis.

In order to compute the χ^2 and F-versions of the test we set $\hat{\tilde{z}}_{it} = (\tilde{x}_{it}', \tilde{\omega}_{it}'\,(\gamma_1, \mathbf{c}_1))'$ and $\tilde{v}_{it} = (\widetilde{x_{it}'q_{it}}, \ldots, \widetilde{x_{it}'q_{it}^m})'$ where $\tilde{w}_{it}(.) = x_{it}g(q_{it}, \gamma, \mathbf{c}) - \sum_t x_{it}g(q_{it}, \gamma, \mathbf{c})/T$ and $\widetilde{x_{it}q_{it}^j} = x_{it}q_{it}^j - \sum_{t=1}^T x_{it}q_{it}^j/T$, $j = 1, \ldots, m$. The test can be computed in three stages as follows:

1. Estimate the PSTR model (2.1) and compute the residual sum of squares SSR_0.

2. Regress $\tilde{y}_{it}$ on $\hat{\tilde{z}}_{it}$ and $\tilde{v}_{it}$, and compute the residual sum of squares SSR_1.

3. Compute the χ^2 and F-versions of the test as follows:

$$\mathrm{LM} = TN(\mathrm{SSR}_0 - \mathrm{SSR}_1)/\mathrm{SSR}_0$$
$$\mathrm{LM_F} = \{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/mk\} \,/\, \{\mathrm{SSR}_1/(TN - N - mk)\}$$

   Statistic LM has an asymptotic χ^2_{mk} distribution under H_0, whereas $\mathrm{LM_F}$ is approximately $F[mk, TN - N - 2mk]$, when H_0' holds.

This testing procedure can be used to determine the number of regimes in the general additive PSTR model. The selection can be done by using the following sequence of hypothesis: Given an estimated PSTR model with $r = r^*$, test the null hypothesis $\mathrm{H}_0 : r = r^*$ against $\mathrm{H}_1 : r = r^* + 1$. If H_0 is not rejected, the testing procedure ends. Otherwise, the null hypothesis $\mathrm{H}_0 : r = r^* + 1$ is tested against the model with $r = r^* + 2$. The testing procedure continues until the first acceptance of H_0.

The sequence of tests for specifying a general additive PSTR model can be carried out as follows:

1. Estimate the linear model and test linearity at significance level α.

2. If linearity is rejected, estimate a single transition PSTR model.

3. Test the hypothesis of no remaining nonlinearity for this model. If the hypothesis of no remaining nonlinearity is rejected at significance level $\tau\alpha$, $0 < \tau < 1$, estimate a double-transition model. The significance level is reduced by a factor τ in order to favour parsimony. Estimation of this model can be carried out in two stages. First, use a grid search of (γ_2, c_2) to find the initial values and then estimate the model by NLS. The grid search is conditional on the estimated values $(\hat{\gamma}_1, \hat{\mathbf{c}}_1')'$ from the previous stage of the process.

4. Continue until the first acceptance of the hypothesis of no remaining nonlinearity.

Eitrheim and Teräsvirta (1996) pointed out potential numerical problems in the computation of the misspecification tests. In particular, when the estimates of γ_j are relatively large, such that the transition between regimes is rapid, the partial derivatives of the transition functions $g_j\left(q_{it}; \hat{\gamma}_j, \hat{\mathbf{c}}_j\right)$, $j = 1, \ldots, r$, with respect to $\left(\hat{\gamma}_j, \hat{\mathbf{c}}_j'\right)'$ obtain about the same value with few

exceptions. As a result, the moment matrix of the auxiliary regression becomes near-singular. However, the contribution of the terms involving these partial derivatives to the test statistic is negligible at large values of γ_j. They can simply be omitted from the auxiliary regression without influencing the empirical size (and power) of the test statistic.

2.4 Simulation study

In this section we investigate small-sample properties of the specification strategy for the PSTR model by simulation. The section is divided in three subsections. The first subsection deals with small sample properties of the NLS estimators. In the second subsection we investigate the size and power properties of the tests of linearity, parameter constancy and no remaining nonlinearity. The final subsection contains results on sequential PSTR model selection.

The design of the Monte Carlo experiment is the following. The number of replications equals 1000. Every experiment is carried out for the following sample sizes: $T = \{5, 10, 20, 40\}$, whereas the number of individuals $N = \{10, 20, 40, 80, 160, 320\}$. The vector of exogenous variables $(x'_{it}, q_{it})'$ is generated independently for each individual following the VAR(1) model

$$\begin{bmatrix} x_{it} \\ q_{it} \end{bmatrix} = \begin{bmatrix} \Theta & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} x_{it-1} \\ q_{it-1} \end{bmatrix} + \begin{bmatrix} \epsilon^x_{it} \\ \epsilon^q_{it} \end{bmatrix}$$

where x_{it} is a (2×1) vector, Θ is a (2×2) matrix, ρ is a scalar, and $(\epsilon^{x\prime}_{it}, \epsilon^q_{it})'$ is a multinormally distributed (3×1) vector with mean zero and positive definite covariance matrix Σ. The first 100 observations for each generated sample are discarded to avoid initialization effects. The endogenous variable is generated from

$$y_{it} = \mu_i + x'_{it}\beta_0 + \sum_{j=1}^{r} g\left(q_{it}; \gamma_j, \mathbf{c}_j\right) x'_{it}\beta_i + e_{it} \tag{2.20}$$

where $g\left(q_{it}; \gamma_j, \mathbf{c}_j\right)$ is defined as in (2.2), $\mu_i = \sigma_\mu u_i$ where u_i is uniform $(0, 1)$, σ_μ is a scalar, and e_{it} is a standard normal variable. The values of r, $(\beta'_0, \ldots, \beta'_r)'$, m, and $(\gamma_j, \mathbf{c}'_j)'$ vary from one design to another. Values of the parameters defining the different designs can be found in Appendix C.

2.4.1 Estimation of the PSTR model

In order to investigate small-sample properties of our estimation algorithm we carry out two experiments. In the first one, we simulate model (2.20) with

$r = 1$ and $m = 1$. We set the value of location vector equal $c = \mathsf{E}(q_i) = 3.5$ where $\mathsf{E}(q_i)$ is the unconditional mean of the transition variable. In the second experiment we generate the data from model (2.20) with $r = 1$ and $m = 2$, with location vector $\mathbf{c} = (3, 4)'$. Finally, to find out the effect of γ on the estimation we consider values $\gamma = 4$, 50. In all cases the estimation is carried out as described in Section 2.3.3. The results for the experiments where $\gamma = 4$

Table 2.1: Mean squared error in the estimation of PSTR models, ($\gamma = 4$)

		$r = 1, m = 1$					$r = 1, m = 2$				
		Mean squared error			Statistics for $\hat{\gamma}$		Mean squared error			Statistics for $\hat{\gamma}$	
N	T	$\hat{\beta}$	$\hat{\gamma}$	$\hat{c}$	Mean	Std	$\hat{\gamma}$	$\hat{\gamma}$	c	Mean	Std
10	5	12.572	0.063	0.090	4.009	0.252	95.980	0.340	4.778	4.008	0.584
	10	4.493	0.029	0.037	4.010	0.170	36.695	0.143	1.420	4.021	0.378
	20	1.994	0.013	0.011	3.994	0.115	14.992	0.063	0.547	3.988	0.251
	40	0.819	0.007	0.007	4.007	0.083	6.963	0.031	0.237	4.005	0.175
20	5	5.489	0.036	0.036	4.025	0.187	46.867	0.162	2.066	3.997	0.403
	10	1.810	0.013	0.014	4.009	0.115	14.033	0.059	0.494	4.005	0.244
	20	0.922	0.005	0.007	4.007	0.074	7.218	0.030	0.255	3.991	0.173
	40	0.370	0.003	0.004	4.002	0.053	3.133	0.014	0.108	3.995	0.116
40	5	2.268	0.014	0.017	4.009	0.117	21.095	0.090	0.778	3.999	0.301
	10	0.919	0.007	0.009	4.003	0.082	8.376	0.037	0.302	4.019	0.191
	20	0.457	0.003	0.003	4.004	0.052	3.332	0.014	0.114	4.004	0.120
	40	0.180	0.001	0.002	3.999	0.039	1.622	0.007	0.055	4.001	0.084
80	5	1.003	0.007	0.009	4.007	0.084	8.932	0.039	0.307	4.003	0.197
	10	0.477	0.003	0.004	4.004	0.059	3.755	0.016	0.126	4.006	0.126
	20	0.250	0.002	0.002	4.000	0.042	1.590	0.007	0.050	4.003	0.085
	40	0.103	0.001	0.001	4.000	0.028	0.789	0.004	0.026	4.002	0.061
160	5	0.576	0.003	0.004	3.997	0.058	4.320	0.018	0.145	3.997	0.135
	10	0.208	0.002	0.002	4.000	0.039	1.977	0.008	0.068	3.995	0.090
	20	0.099	0.001	0.001	4.000	0.026	0.733	0.003	0.025	4.002	0.057
	40	0.049	0.000	0.000	3.998	0.019	0.449	0.002	0.016	4.001	0.045
320	5	0.290	0.002	0.002	4.001	0.046	1.745	0.007	0.056	4.005	0.083
	10	0.116	0.001	0.001	4.001	0.030	0.918	0.004	0.030	4.002	0.066
	20	0.052	0.000	0.000	3.998	0.018	0.395	0.002	2.018	4.003	0.042
	40	0.022	0.000	0.000	3.999	0.014	0.208	0.001	0.007	3.999	0.030

are in Table 2.1 and the ones where $\gamma = 50$ in Table 2.2. The reported statistics are the mean squared error (MSE) in estimating $(\beta_0', \beta_1')'$, γ and $\mathbf{c}$ and the Monte Carlo mean and standard deviation of $\hat{\gamma}$. The tables are divided in two panels whose left-hand sides contain the results for $m = 1$, whereas the results of the second experiment with $m = 2$ can be found in the right-hand panel. The results indicate that the parameter estimates $(\beta_0', \beta_1')'$ and $\hat{c}$ of the PSTR model can be estimated with reasonable accuracy. However, by

comparing the MSEs of $(\beta_0', \beta_1')'$, and $\mathbf{c}$ between Table 2.1 and Table 2.2, it is seen that the accuracy is higher when γ is large than when it is small. In particular, for $\gamma = 4$ and $m = 1$ the average MSE in estimating β and $\mathbf{c}$ equals 1.481 and 0.011, respectively, whereas for $\gamma = 50$ the average MSE for the same parameters are 0.896 and 0.003. This may be due to the fact that for $\gamma = \infty$ the elements of $\mathbf{c}$ are estimated superconsistently. Consequently, when γ is finite but sufficiently large, $\mathbf{c}$ is estimated more accurately than is the case when γ is small. This result contrasts with the fact that small values of γ can be estimated with greater accuracy than large values. For example, for $N = 10$ and $T = 5$, the MSE for $\gamma = 4$ equals 0.063 while the MSE for $\gamma = 50$ is 13000. But then, when γ is large the magnitude of the MSE does not play a role because as long as the estimate is also large the estimation error does not matter very much. At the other end, if γ is close to zero estimation may be difficult because then the model is close to being unidentified and the sequence of estimates may not converge.

Finally, by comparing the MSE across values of N for given T, one can notice the gains from the panel in estimating γ. For instance, when $m = 1$, $T = 10$ and $\gamma = 4$ the MSE for $N = 10, 40, 320$ equal 4.493, 0.919 and 0.052, respectively. This is not an unexpected outcome because the accuracy of the estimate should increase with the number of observed transitions.

2.4.2 Small-sample properties of the misspecification tests

In this section we investigate the small-sample properties of our misspecification tests. In particular, we investigate the size and power of these tests for different values of N and T. The section is divided in three subsections and each subsection contains the results for a particular test.

Linearity test

Previous studies have documented that the F-version of the test has better size properties in small samples than the asymptotic χ^2-based statistic. For this reason, we only report the results for the F-version. [See Eitrheim and Teräsvirta (1996) and van Dijk, Teräsvirta, and Franses (2002) for details]. Additionally, since the auxiliary regression (2.4) with $m = 3$ has power against the alternative model (2.1) with $m = 1$ and 2 in (2.2), we compute the test statistic using $m = 1, 2, 3$ in (2.4)[3]. We denote by m^* the maximum power of

[3]The test based on the auxiliary regression (2.4) assuming $m^* = 3$ has power against the alternative models with $m = 1$ and $m = 2$ because it can be interpreted as having been

q_{it} in the auxiliary regression (2.4). The order m^* of the auxiliary regression (2.4) may affects both the size and power of the test. For instance, approximations to the alternative model based on large values of m^* imply a loss of degrees of freedom which may cause problems, but they may add power to the test because the approximation to the nonlinear component improves with increasing m^*. In order to investigate the empirical size of our linearity test we generate 1000 samples of a linear panel with fixed effects. The results are based on the nominal significance level 0.05 and can be found in Table 2.3. Each column contains results for the test statistic based on one auxiliary regression with m^* =1, 2, 3. It is seen that the empirical size of the test is close to the nominal size at all sample sizes. The loss of degrees of freedom associated with large values of m^* compared to small ones does not seem to affect the size of the test.

To investigate the power of the test we generate samples from the PSTR model (2.20) with $r = 1$ and for $m = 1, 2$. The model under the alternative is thus a standard PSTR model with either a monotonically increasing ($m = 1$) or symmetric ($m = 2$) transition function. To find out the extent to which the value of γ affects the power we set $\gamma = 2$ and $\gamma = 5$. The values of the remaining parameters can be found is in Appendix C. Table 2.4 contains the estimated powers of the test. The left-hand panel shows the power for the PSTR model with $m = 1$. The results for $m = 2$ appear in the right-hand panel. In all cases, we report the power of the test based on three auxiliary regressions (2.4) with $m^* = 1, 2, 3$. It is seen that the test based on the auxiliary regression with $m^* = m$ has best power for both alternative models $m = 1$ and $m = 2$. Moreover, from the right-hand panel it appears that when $m = 2$ the test based on $m^* = 1$ does not have power, while the test based on $m^* = 3$ has only little less power than the test based on $m^* = 2$. The results from this section indicate, not unexpectedly, that the test based on the auxiliary regression with $m^* = 2$ is preferable to the others, especially in situations where the transition is symmetric around $(c_1 + c_2)/2$ ($m = 2$).

Parameter constancy test

Next we investigate the size and power properties of a parameter constancy test. To estimate the size of test we generate the replications with the PSTR

derived by replacing the transition (2.2) in (2.1) by its second order Taylor expansion. When $m = 1$ in (2.2) the second order Taylor expansion yields an auxiliary regression with $m^* = 2$, and when $m = 2$ the second order Taylor expansion of transition function yields an auxiliary regression with $m^* = 4$. Consequently, the test based on $m^* = 3$ has power against both alternative models, $m = 1$ and $m = 2$.

model (2.20) with constant parameters, setting $r = 1$, $m = 1$, $\gamma_1 = 3$ and $c_1 = 3.5$. The values of the remaining parameters are reported in Appendix C. For easy of presentation we denote by h^* the maximum power of (t/T) in the auxiliary regression (2.14). As before, the auxiliary regression (2.14) with $h^* = 3$ has power against alternative models with $h = 1$ and $h = 2$ because it nests the ones based on $h^* = 1$ and $h^* = 2$. The power of the test is computed by simulating the model under the alternative hypothesis (2.11). We consider two alternative models. In the first one, we allow one transition in the parameters in the middle of the sample: $h = 1$ and $c_2 = 0.5$. In the second one, the transition is symmetric around $(c_{21} + c_{22})/2 = 0.5$ with $c_{21} = 0.3$ and $c_{22} = 0.7$. In both cases $\gamma_2 = 2$ in (2.13). The values of the remaining parameters can again be found in Appendix C. We assume that the transition occurs at the same time for all N individuals and that the change in the parameters is the same for all of them.

Table 2.5, with three panels, contains the estimated size and power of the test. The leftmost panel contains the empirical size of the test, the other two have the power results. All three panels are divided in three columns, one for each value of h^*. The results indicate that the F-version of the test has the correct size for all sample sizes and values of h^*. Furthermore, the test based on $h^* = 3$ has good power against the alternative models with $h = 1$ and $h = 2$. This suggests that in terms of power, one is bound to gain more by using a closer approximation to the model under the alternative than what one loses by giving up degrees of freedom compared to cases $h^* = 1$ and $h^* = 2$.

Test of no remaining nonlinearity and properties of the procedure for determining the number of regimes in a multiple regime PSTR panel

The size simulations of the test of no remaining nonlinearity are carried out using (2.20) and setting $r = 1$, $m = 1$, $\gamma_1 = 3$ and $c_1 = 3.5$. In order to estimate the power of the test we generate the samples from the same model with $r = 2$. The parameters in the first transition function are the same as in the size simulations. The second transition function is of type (2.2) with $m = 1$ and $c = 2$. To find out the effect of γ_2 in the power of the test we simulate the test for $\gamma_2 = 2, 5$. The remaining parameters are reported in the Appendix C. As before, we consider only the F-version of the test and use three different auxiliary regressions (2.19) with $m^* = 1, 2, 3$. The estimated size and power at the 5% nominal level are shown in Table 2.6. The results indicate that the test based on the auxiliary regression with $m^* = 2$ has the best properties. First, its size is close to the nominal size and second, it has

better power than the other alternatives.

Finally, we consider the performance of our procedure for determining the number of regimes discussed in Section 2.3.4. The simulations are carried out using model (2.20) with two and three transitions, that is, $r = 2$ and $r = 3$. The transition functions are defined as in (2.2) with $m = 1$. The remaining parameters in the transition functions are: The first transition function: $\gamma_1 = 5$, $c_1 = 2.869$; the second transition function $\gamma_2 = 10$, $c_2 = 4.35$, and the third transition function $\gamma_3 = 20$, $c_3 = 3.47$. The remaining parameters can be found in Appendix C. We set $\tau = 0.5$, which means that we reduce the significance level of the test by one half at each step of the sequence. We use the F-version of the test of no remaining nonlinearity that is based on the auxiliary regression (2.19) with $m^* = 2$. The frequency at which the procedure selects a particular value of r from $r = 1$ to 4 appears in Table 2.7. The right-hand panel in Table 2.7 contains results for $r = 2$ and the left-hand one those for $r = 3$. They show that, for a given number of observations, it is easier to identify a double-transition PSTR model that a three-regime PSTR model, which might be not surprising. In fact, correct identification of a three-regime model requires relatively large sample sizes. The procedure seems to work relatively well for sample panels with $T = 5$, 10, when $N \geq 40$. An interesting detail is that "crossing" to three transitions through two is difficult in small samples.

Table 2.2: Mean squared error in the estimation of PSTR models, ($\gamma = 50$)

		$r = 1, m = 1$					$r = 1, m = 2$				
		Mean squared error			Statistics for $\hat{\gamma}$		Mean squared error			Statistics for $\hat{\gamma}$	
N	T	$\hat{\beta}$	$\hat{\gamma}$	$\hat{c}$	Mean	Std	$\hat{\beta}$	$\hat{\gamma}$	$\hat{c}$	Mean	Std
10	5	7.388	1.32×10^{4}	5.52×10^{-2}	92.401	360.9	7.564	8.40×10^{2}	2.30×10^{-0}	66.599	90.29
	10	2.661	5.14×10^{2}	7.83×10^{-3}	58.113	71.37	2.938	35.505	9.35×10^{-3}	51.326	5.821
	20	1.090	13.535	7.37×10^{-4}	50.031	3.686	1.229	7.506	3.69×10^{-3}	50.523	2.695
	40	0.491	5.690	2.99×10^{-4}	50.293	2.372	0.532	3.206	6.68×10^{-4}	49.891	1.791
20	5	3.174	1.70×10^{3}	1.55×10^{-2}	66.703	129.4	3.039	8.94×10^{1}	1.73×10^{-2}	52.905	29.81
	10	1.247	14.125	7.51×10^{-4}	50.173	3.762	1.321	8.998	1.76×10^{-3}	50.344	2.986
	20	0.540	4.941	2.60×10^{-4}	50.113	2.224	0.600	2.943	6.67×10^{-4}	50.170	1.711
	40	0.299	2.568	1.38×10^{-4}	50.052	1.605	0.320	1.331	3.37×10^{-4}	49.964	1.156
40	5	1.545	13.545	1.00×10^{-3}	50.092	3.687	1.505	8.296	2.02×10^{-3}	50.165	2.881
	10	0.575	6.194	2.88×10^{-4}	50.384	2.464	0.625	3.848	6.86×10^{-4}	49.988	1.966
	20	0.248	2.693	1.26×10^{-4}	50.068	1.643	0.309	1.215	2.89×10^{-4}	50.025	1.104
	40	0.118	1.067	5.85×10^{-5}	50.032	1.035	0.123	0.660	1.41×10^{-4}	50.061	0.811
80	5	0.757	6.143	3.24×10^{-4}	50.096	2.482	0.786	3.053	6.74×10^{-4}	50.035	1.750
	10	0.286	2.518	1.33×10^{-4}	50.136	1.584	0.292	1.571	3.04×10^{-4}	50.031	1.255
	20	0.133	1.253	6.48×10^{-5}	50.077	1.119	0.136	0.801	1.32×10^{-4}	49.976	0.896
	40	0.066	0.577	2.78×10^{-5}	49.990	0.761	0.065	0.334	7.37×10^{-5}	49.988	0.579
160	5	0.358	3.235	1.26×10^{-4}	50.311	1.775	0.340	1.628	4.07×10^{-4}	50.062	1.277
	10	0.141	1.209	6.80×10^{-5}	50.026	1.101	0.128	0.788	1.41×10^{-4}	49.984	0.890
	20	0.058	0.528	3.15×10^{-5}	50.042	0.727	0.071	0.355	6.99×10^{-5}	50.016	0.596
	40	0.033	0.295	1.46×10^{-5}	49.956	0.543	0.035	0.165	3.76×10^{-5}	49.999	0.407
320	5	0.164	1.483	6.28×10^{-5}	50.060	1.219	0.172	0.799	1.70×10^{-4}	50.019	0.895
	10	0.077	0.682	3.46×10^{-5}	49.967	0.827	0.077	0.352	7.61×10^{-5}	49.987	0.594
	20	0.033	0.279	1.53×10^{-5}	49.977	0.529	0.035	0.158	3.53×10^{-5}	50.007	0.398
	40	0.013	0.154	7.48×10^{-6}	50.026	0.393	0.018	0.098	1.51×10^{-5}	50.010	0.314

Table 2.3: Empirical size of the linearity test at 0.05 nominal level

N	T	$m^* = 1$	$m^* = 2$	$m^* = 3$
10	5	0.049	0.047	0.049
	10	0.050	0.048	0.046
	20	0.048	0.052	0.046
	40	0.057	0.051	0.047
20	5	0.049	0.052	0.049
	10	0.051	0.043	0.043
	20	0.044	0.047	0.048
	40	0.049	0.059	0.050
40	5	0.043	0.045	0.042
	10	0.040	0.043	0.044
	20	0.051	0.046	0.043
	40	0.056	0.044	0.042
80	5	0.053	0.053	0.052
	10	0.052	0.049	0.049
	20	0.052	0.049	0.047
	40	0.051	0.045	0.047
160	5	0.050	0.059	0.051
	10	0.056	0.059	0.055
	20	0.056	0.059	0.051
	40	0.052	0.045	0.037
320	5	0.048	0.046	0.043
	10	0.053	0.049	0.048
	20	0.044	0.048	0.050
	40	0.046	0.053	0.047

The m* in the name of the columns indicates the maximum value of m in the auxiliary regression. The nominal size is 0.05

Table 2.4: Empirical power of the linearity test at the 0.05 level of significance: m as in (2.2), m^* equals the value of m in the auxiliary regression

		$m=1$						$m=2$					
		$\gamma=2$			$\gamma=5$			$\gamma=2$			$\gamma=5$		
N	T	$m^*=1$	$m^*=2$	$m^*=3$	$m^*=1$	$m^*=2$	$m^*=3$	$m^*=1$	$m^*=2$	$m^*=3$	$m^*=1$	$m^*=2$	$m^*=3$
10	5	0.098	0.069	0.064	0.085	0.075	0.066	0.055	0.045	0.055	0.059	0.086	0.065
	10	0.141	0.117	0.097	0.190	0.168	0.139	0.054	0.083	0.073	0.046	0.115	0.096
	20	0.244	0.192	0.157	0.396	0.359	0.302	0.049	0.122	0.108	0.053	0.261	0.231
	40	0.459	0.370	0.332	0.709	0.699	0.649	0.058	0.218	0.180	0.049	0.518	0.464
20	5	0.105	0.090	0.079	0.167	0.154	0.138	0.046	0.067	0.055	0.054	0.120	0.111
	10	0.224	0.188	0.146	0.370	0.338	0.285	0.053	0.120	0.088	0.058	0.228	0.203
	20	0.496	0.409	0.349	0.712	0.708	0.644	0.051	0.210	0.176	0.068	0.482	0.440
	40	0.802	0.724	0.648	0.956	0.956	0.940	0.046	0.442	0.393	0.063	0.850	0.815
40	5	0.200	0.164	0.122	0.305	0.262	0.230	0.054	0.112	0.096	0.046	0.210	0.173
	10	0.418	0.334	0.285	0.647	0.620	0.543	0.052	0.214	0.183	0.053	0.450	0.417
	20	0.746	0.661	0.600	0.951	0.950	0.930	0.061	0.430	0.376	0.059	0.822	0.784
	40	0.980	0.962	0.942	1.000	0.999	1.000	0.052	0.764	0.704	0.056	0.991	0.985
80	5	0.368	0.292	0.237	0.591	0.576	0.493	0.050	0.177	0.142	0.046	0.382	0.343
	10	0.714	0.617	0.561	0.925	0.928	0.888	0.046	0.388	0.336	0.052	0.778	0.741
	20	0.975	0.965	0.941	0.999	1.000	0.998	0.048	0.767	0.721	0.061	0.993	0.984
	40	1.000	1.000	0.999	1.000	1.000	1.000	0.054	0.972	0.957	0.056	1.000	1.000
160	5	0.642	0.566	0.491	0.875	0.882	0.818	0.048	0.362	0.314	0.064	0.717	0.665
	10	0.962	0.929	0.904	0.998	0.999	0.999	0.053	0.713	0.667	0.067	0.984	0.974
	20	1.000	1.000	1.000	1.000	1.000	1.000	0.053	0.979	0.965	0.067	1.000	1.000
	40	1.000	1.000	1.000	1.000	1.000	1.000	0.056	1.000	1.000	0.077	1.000	1.000
320	5	0.938	0.889	0.845	0.995	0.999	0.999	0.071	0.663	0.607	0.069	0.968	0.956
	10	1.000	1.000	0.999	1.000	1.000	1.000	0.047	0.967	0.954	0.059	1.000	1.000
	20	1.000	1.000	1.000	1.000	1.000	1.000	0.051	1.000	1.000	0.059	1.000	1.000
	40	1.000	1.000	1.000	1.000	1.000	1.000	0.063	1.000	1.000	0.096	1.000	1.000

Table 2.5: Empirical size and power of the parameter constancy test at the 0.05 level of significance

		Size			Power					
					$h=1$			$h=2$		
N	T	$h^*=1$	$=2$	$=3$	$=1$	$=2$	$=3$	$=1$	$=2$	$=3$
10	5	0.046	0.043	0.046	0.091	0.073	0.056	0.051	0.088	0.079
	10	0.056	0.041	0.036	0.208	0.162	0.146	0.048	0.094	0.084
	20	0.052	0.051	0.046	0.472	0.329	0.328	0.068	0.156	0.208
	40	0.047	0.049	0.043	0.828	0.713	0.714	0.099	0.334	0.530
20	5	0.042	0.054	0.049	0.162	0.119	0.107	0.053	0.193	0.188
	10	0.046	0.048	0.049	0.444	0.343	0.308	0.067	0.175	0.219
	20	0.041	0.048	0.043	0.811	0.711	0.682	0.076	0.267	0.418
	40	0.052	0.045	0.044	0.996	0.983	0.984	0.199	0.625	0.868
40	5	0.056	0.049	0.046	0.338	0.254	0.211	0.060	0.430	0.425
	10	0.046	0.048	0.045	0.793	0.675	0.610	0.086	0.365	0.473
	20	0.051	0.046	0.043	0.984	0.963	0.970	0.181	0.636	0.832
	40	0.052	0.046	0.041	1.000	1.000	1.000	0.426	0.954	0.998
80	5	0.049	0.045	0.047	0.675	0.568	0.487	0.084	0.801	0.793
	10	0.055	0.048	0.054	0.986	0.970	0.956	0.127	0.749	0.873
	20	0.049	0.050	0.052	1.000	1.000	1.000	0.300	0.926	0.994
	40	0.050	0.052	0.054	1.000	1.000	1.000	0.750	0.999	1.000
160	5	0.042	0.056	0.051	0.940	0.895	0.852	0.102	0.992	0.995
	10	0.049	0.055	0.051	1.000	1.000	1.000	0.256	0.981	0.998
	20	0.052	0.051	0.053	1.000	1.000	1.000	0.629	1.000	1.000
	40	0.051	0.049	0.052	1.000	1.000	1.000	0.980	1.000	1.000
320	5	0.050	0.049	0.049	1.000	0.998	0.992	0.198	1.000	1.000
	10	0.050	0.049	0.056	1.000	1.000	1.000	0.485	1.000	1.000
	20	0.039	0.043	0.044	1.000	1.000	1.000	0.910	1.000	1.000
	40	0.049	0.049	0.048	1.000	1.000	1.000	1.000	1.000	1.000

The left-hand panel contains the estimated size while power results are presented in the two right hand side panels. The panel with the power results is divided in two three-column panels. The first panel shows the estimated power of the test when the alternative model is (2.11) with $h = 1$ in (2.13). The other power panel displays the estimated power of the test when the alternative model is (2.11) with $h = 2$ in (2.13). The size and power calculation are based on the 5% critical values of the F distribution. The h^* in the column names indicates the maximum value of h in the auxiliary regression (2.14).

Table 2.6: Empirical size and power of the test for no remaining nonlinearity at the 0.05 level of significance

		Size			Power					
					$m=1$			$m=2$		
N	T	$m^*=1$	$=2$	$=3$	$=1$	$=2$	$=3$	$=1$	$=2$	$=3$
10	5	0.044	0.037	0.038	0.042	0.051	0.053	0.042	0.052	0.057
	10	0.056	0.045	0.050	0.053	0.054	0.041	0.075	0.070	0.058
	20	0.035	0.048	0.046	0.056	0.065	0.063	0.094	0.130	0.107
	40	0.042	0.038	0.034	0.086	0.098	0.072	0.140	0.229	0.216
20	5	0.045	0.052	0.041	0.053	0.053	0.038	0.058	0.062	0.058
	10	0.042	0.053	0.046	0.063	0.052	0.048	0.081	0.105	0.096
	20	0.054	0.040	0.049	0.062	0.104	0.082	0.141	0.239	0.195
	40	0.042	0.046	0.038	0.095	0.170	0.155	0.189	0.432	0.376
40	5	0.041	0.040	0.031	0.064	0.064	0.060	0.082	0.123	0.111
	10	0.049	0.047	0.045	0.076	0.097	0.086	0.129	0.219	0.191
	20	0.051	0.044	0.040	0.082	0.143	0.126	0.194	0.444	0.390
	40	0.054	0.052	0.040	0.141	0.294	0.237	0.377	0.784	0.745
80	5	0.055	0.038	0.035	0.057	0.082	0.086	0.122	0.185	0.162
	10	0.049	0.047	0.043	0.088	0.135	0.126	0.201	0.401	0.350
	20	0.051	0.035	0.037	0.123	0.269	0.221	0.354	0.754	0.712
	40	0.042	0.044	0.044	0.244	0.575	0.503	0.644	0.981	0.966
160	5	0.049	0.044	0.042	0.083	0.132	0.112	0.174	0.367	0.316
	10	0.049	0.051	0.052	0.137	0.266	0.222	0.327	0.705	0.668
	20	0.048	0.053	0.046	0.249	0.544	0.464	0.640	0.979	0.968
	40	0.049	0.061	0.051	0.438	0.885	0.848	0.922	1.000	1.000
320	5	0.051	0.059	0.058	0.111	0.228	0.180	0.320	0.662	0.626
	10	0.048	0.054	0.047	0.209	0.517	0.455	0.632	0.962	0.952
	20	0.053	0.055	0.057	0.408	0.847	0.801	0.914	1.000	1.000
	40	0.046	0.052	0.055	0.735	0.996	0.993	1.000	1.000	1.000

The m^* in the name of the columns indicates the maximum value of m in the auxiliary regression (2.19).

Table 2.7: Small sample behaviour of the algorithm for determining the number of regimes in a Multiple PSTR model

		True $r = 2$			True $r = 3$			
N	T	$r = 1$	$r = 2$	$r > 2$	$r = 1$	$r = 2$	$r = 3$	$r > 3$
10	5	0.015	0.961	0.024	0.792	0.086	0.110	0.012
	10	0.000	0.978	0.022	0.690	0.004	0.290	0.016
	20	0.000	0.972	0.028	0.436	0.000	0.556	0.008
	40	0.000	0.971	0.029	0.116	0.000	0.880	0.004
20	5	0.000	0.985	0.015	0.716	0.020	0.252	0.012
	10	0.000	0.978	0.022	0.500	0.000	0.492	0.008
	20	0.000	0.980	0.020	0.138	0.000	0.850	0.012
	40	0.000	0.984	0.016	0.000	0.000	0.980	0.020
40	5	0.000	0.977	0.023	0.586	0.000	0.404	0.010
	10	0.000	0.985	0.015	0.164	0.000	0.828	0.008
	20	0.000	0.981	0.019	0.002	0.000	0.986	0.012
	40	0.000	0.979	0.021	0.000	0.000	0.980	0.020
80	5	0.000	0.983	0.017	0.260	0.000	0.722	0.018
	10	0.000	0.976	0.024	0.004	0.000	0.988	0.008
	20	0.000	0.980	0.020	0.000	0.000	1.000	0.000
	40	0.000	0.973	0.027	0.000	0.000	0.994	0.006
160	5	0.000	0.977	0.023	0.014	0.000	0.978	0.008
	10	0.000	0.975	0.025	0.000	0.000	0.986	0.014
	20	0.000	0.983	0.017	0.000	0.000	0.984	0.016
	40	0.000	0.984	0.016	0.000	0.000	0.990	0.010
320	5	0.000	0.976	0.024	0.000	0.000	0.988	0.012
	10	0.000	0.969	0.031	0.000	0.000	0.988	0.012
	20	0.000	0.972	0.028	0.000	0.000	0.988	0.012
	40	0.000	0.983	0.017	0.000	0.000	0.988	0.012

Each column contains the proportion of times the algorithm selects a given r. The left-hand panel contains results for the two-regime model while the right-hand panel the results for $r = 3$. The nominal level of the test is 0.05. We half the nominal level of the test at each step of the algorithm; $\tau = 0.5$.

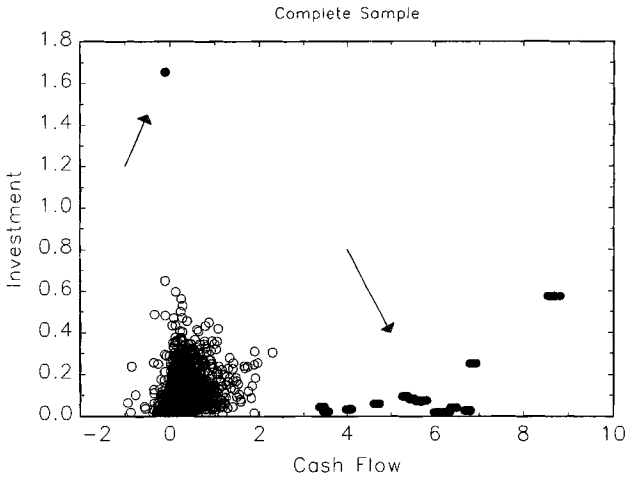
2.5 An application: Investment and financial constraints

In this section we present an application of the modelling cycle for PSTR models. We choose the same economic problem and data set as Hansen (1999a)[4] and compare our results to the ones in that article. The economic question under investigation is whether firms that are financially constrained behave differently from financially unconstrained firms when it comes to financing investment. Fazzari, Hubbard, and Petersen (1988) argued that when the capital market is imperfect, the firm's financial structure is not irrelevant to its investment decisions because internal and external funds are not perfect substitutes. In their view, when a firm is facing credit constraints, investment may depend upon the availability of internal funds, such as the firm's cash flow. For instance, in presence of asymmetric information between the firm and the provider of external finance, the cost of external funding is higher than the cost of internal funding, which favours the use of internal resources to finance investment. Furthermore, when the debt level is high the firm is likely to be financially constrained and consequently, its cash flow will be positively correlated with the investment rate.

This situation calls for a PSTR model such that the transition variable measures the financial position of each firm and period. A natural candidate, and the one Hansen (1999a) used, is the debt-to-asset ratio. It measures the existing debt level and can be treated as a proxy for availability of external funds to the firm. This is because external providers of financing may be reluctant to lend capital to strongly indebted firms. The data set is extracted from the one used by Hall and Hall (1993) and consists of 565 firms observed from 1973 to 1987. The variables include the ratio of investment to capital I_{it}, the ratio of total market value to assets Q_{it}, the ratio of long-term debt to assets D_{it} and the ratio of cash flow to assets CF_{it}. We delete two firms from the original sample because they have atypical investment-cash flow relationship. Figure 2.1 shows the scatter plot of investment *versus* cash flow. The observations for the deleted firms are represented by solid circles. As is seen, these firms have either extremely large levels of cash flow with low levels of investment or extremely large levels of investment with very low levels of cash flow.

[4]The data set is available in Bruce Hansen web page `http://www.ssc.wisc.edu/~bhansen/`

Figure 2.1: Scatter plot of the cash flow and investment pairs for the complete sample



An application of the PSTR modelling strategy Our starting point is that investment can be characterized by a Multiple Regime PSTR model (2.16). Our maintained model for investment thus has the form

$$I_{it} = \mu_i + \theta_1 Q_{it-1} + \theta_2 Q_{it-1}^2 + \theta_3 Q_{it-1}^3 + \theta_4 D_{it-1}$$
$$+\theta_5 Q_{it-1} D_{it-1} + \beta_0 CF_{it-1} + \sum_{j=1}^{r} \beta_j CF_{it-1} g\left(D_{it-1}; \theta_j\right) + e_{it} \qquad (2.21)$$

where the transition functions $g\left(D_{it-1}; \theta_j\right)$ $j = 1, \ldots, r$ are of type (2.2) with $m = 1$. Following Hansen (1999a) we include the terms Q_{it-1}^j for $j = 2, 3$, D_{it-1} and $Q_{it-1} D_{it-1}$ to account for possible omitted variables.

The multiple-regime PSTR model is specified as suggested in Section 2.3.4. The results of the linearity test are presented in Table 2.8. Since linearity is rejected at 5% significance level we estimate a Multiple PSTR model with $r = 1$ and $m = 1$ and test whether another transition is required. The next null hypothesis of $r = 1$ is not rejected, so our final model has a single transition.

Table 2.9 contains the parameter estimates of the single transition model. It is seen from the table and Figure 2.2 that the transition is quite sharp and the model thus is close to a two-regime PTR model. The first regime contains the firms with low debt levels, $D_{it-1} < 0.0154$ and the second regime the more strongly indebted firms, $D_{it-1} \geq 0.0154$.

The combined "parameter" of the cash flow as a function of the debt level

Table 2.8: Determining the number of regimes using the test for no remaining nonlinearity

		$H_0 : r = 0$ *vs* $H_0 : r = 1$
		Actual Significance level: $\alpha = 0.05$
$F2$	8.58	2×10^{-10}
$F3$	6.00	4×10^{-11}
		$H_0 : r = 1$ *vs* $H_0 : r = 2$
		Actual Significance level: $\alpha = 0.025$
$F2$	2.31	0.10
$F3$	2.27	0.08

We use 0.05 as significance level and assume that $\tau = 0.5$.

Table 2.9: Parameter estimates for the final PSTR model

Regressor	Coefficient estimate	Standard error
Q_{it-1}	0.0118	0.0010
$Q_{it-1}^2/10^3$	-0.2602	0.0365
$Q_{it-1}^3/10^6$	1.4500	0.2700
D_{it-1}	-0.0218	0.0026
$Q_{it-1}D_{it-1}$	0.0017	0.0016
CF_{it-1}	0.0539	0.0054
$CF_{it-1}g\left(D_{it-1};\gamma_1,c_1\right)$	0.0355	0.0047
Transition Functions		
Parameter	Coefficient estimate	Standard error
γ_1	0.508×10^5	139.28
c_1	0.01554	0.0055

equals

$$\beta^{CF}\left(D_{it-1}\right) = \beta_1 + \beta_2 g\left(D_{it-1};\gamma_1,c_1\right)$$

where β_1 and β_2, are the coefficients of the cash flow variable CF_{it-1} in Table 2.9. Figure 2.2 shows the estimated parameter of the cash flow as a function of the debt level. Our results suggest a positive relationship between cash flow and investment which conforms to the prediction of Fazzari, Hubbard, and Petersen (1988).

In order to compare the results of our methodology with those obtained following Hansen's approach we estimate (2.21) as a multiple PTR model. To determine the number of thresholds we apply Hansen's sequential procedure which is similar to ours. The difference is that the candidate models are PTR models. The procedure works as follows. First, estimate a linear model

and test it against a model with one threshold (two regimes). If the null hypothesis is rejected, estimate a single-threshold model and test it against a double-threshold one. The procedure is continued until the hypothesis no additional threshold is not rejected.

Table 2.10 contains the results of the linearity test. They indicate that a two-regime model is enough to characterize the nonlinearity in the data. The final estimated PTR model is reported in Table 2.11. The threshold value $c_1 = 0.0157$, which is very close to the estimated one in the PSTR model. This is expected since the estimate of γ in the PSTR model is large. Figure 2.2 contains the coefficient of the cash flow as a function of the debt level in the estimated PTR model. The resulting graph is similar to the one obtained with the PSTR model.

Table 2.10: Test for threshold effects using Hansen (1999a)

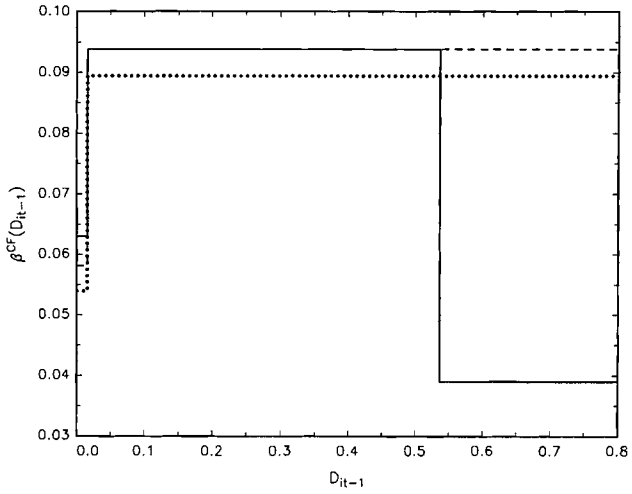
Test a for single threshold	
F_1	44.3
P-value	0.0000
(10%,5%,1% critical values)	(13.9,18.4,25.8)
Test a for double threshold	
F_2	9.1
P-value	0.26
(10%,5%,1% critical values)	(12.4 ,15.5 ,19.7)

Figure 2.2 also contains the coefficient of the cash flow as a function of the debt level in the original PTR model of Hansen (1999a). Hansen used the complete data set and concludes that a three-regime threshold model characterized the data. His results indicated that the relationship between investment and cash flow is nonmonotonic in the sense that the connection between these variables is weaker for very highly indebted firms. He argued that there is considerable uncertainty in the estimate of the cash flow coefficient in the last regime. From our results it becomes clear that this uncertainty is due to the presence of the two outliers in the original sample. When they are removed, our results agree with the ones obtained using the PTR approach.

Table 2.11: Parameter estimates for the two-regime model of Hansen (1999a)

Regressor	Coefficient estimate	Standard errors
Q_{it-1}	0.0117	0.0009
$Q_{it}^2/10^3$	-0.2540	0.0285
$Q_{it}^3/10^6$	1.4028	0.2151
D_{it-1}	-0.0268	0.0046
$Q_{it-1}D_{it-1}$	0.0022	0.0014
$CF_{it-1}I(D_{it-1} \leq c_1)$	0.0582	0.008
$CF_{it-1}I(D_{it-1} \geq c_1)$	0.0938	0.007
Threshold Values		
Parameter	Coefficient estimate	
c_1	0.0157	

Figure 2.2: Estimates of the cash flow parameter as a function of the lagged debt level, D_{it-1}



Dotted line: PSTR.
Dashed line: Panel threshold estimates with restricted sample.
Solid line: Panel threshold estimates with complete sample.

2.6 Conclusions

We introduce a panel model in which the parameters can change as a function of an exogenous variable. The model is thus an alternative to the Panel Threshold model by Hansen (1999a). In the present context however, standard asymptotic theory can be used as the likelihood function in the PSTR model is a continuous function of the parameters. We present tests for linearity, parameter constancy and remaining non linearity. These tests, serve as both specification and misspecification tests. The small sample properties of the proposed statistics was investigated by simulation and the results indicate that they behave well even in panels with small N and T.

Bibliography

AMEMIYA, T. (1985): *Advanced Econometrics.* Hardvard University Press.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

BALTAGI, B. H., AND Q. LI (1995): "Testing AR(1) against MA(1) disturbances in an error component model," *Journal of Econometrics*, 68, 133–151.

CHAMBERLAIN, G. (1984): "Panel data," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2, pp. 1248–1318, Amsterdam:. North-Holland.

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 alternative.," *Biometrika*, 74, 33–43.

EITRHEIM, Ø., AND T. TERÄSVIRTA (1996): "Testing the adequacy of smooth transition autoregressive models," *Journal of Econometrics*, 74, 59–75.

FAZZARI, S., G. HUBBARD, AND B. PETERSEN (1988): "Financial constrains and corporate investments," *Brokings Papers on Economic Activity*, pp. 141–195.

GRANGER, C. W. J., AND T. TERÄSVIRTA (1993): *Modelling nonlinear economic relationships.* Oxford: Oxford University Press.

HALL, B. H., AND R. E. HALL (1993): "The value and performance of U.S. corporations.," *Brookings Papers on Economic Activity*, pp. 1–34.

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

——— (1999a): "Threshold effects in non-dynamic panels: Estimation, testing and inference," *Journal of Econometrics*, 93, 345–368.

——— (1999b): "Testing for linearity," *Journal of Economic Surveys*, 13, 551–576.

——— (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603.

JANSEN, E. S., AND T. TERÄSVIRTA (1996): "Testing parameter constancy and super exogeneity in econometric equations," *Oxford Bulletin of Economics and Statistics*, 58, 735–763.

LEYBOURNE, S., P. NEWBOLD, AND D. VOUGAS (1998): "Unit roots and smooth transitions," *Journal of Time Series Analysis*, 19, 83–97.

LUNDBERGH, S., T. TERÄSVIRTA, AND D. VAN DIJK (2003): "Time-varying smooth transition autoregressive model," *Journal of Business and Economic Statistics*, 21, 104–121.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

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

TERÄSVIRTA, T. (1998): "Modelling economic relationships with smooth transition regressions," in *Handbook of applied economic statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. New York: Marcel Dekker.

VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): "Smooth transition autoregressive models - a survey of recent developments," *Econometric Reviews*, 21, 1–47.

WHITE, H. (1980): "Nonlinear regression on cross-section data," *Econometrica*, 48, 721–746.

——— (2000): *Asymptotic theory for econometricians*. 2nd edition - Academic Press.

Appendix A

Asymptotic properties

This Appendix contains the proof of Theorem 1. The following Lemma is a standard result but it is stated to clarify the notation in this section and Appendix B.

Lemma 1 *Let* $W_i = (w'_{i1}, \ldots, w'_{iT})'$ *be a* $(T \times k)$ *matrix of random variables and let* $Q_T = (c_1, \ldots, c_T) = I_T - \frac{1}{T}\iota\iota'$ *be the within transformation matrix where* ι *is a* $(T \times 1)$ *vector of ones. Then* $W'_i Q'_T Q_T W_i = \sum_{s=1}^{T} \sum_{t=1}^{T} c_{st} w_{it} w'_{is}$ *where* $c_{st} = c'_s c_t$ *is finite non random function of* T.

A.1 Linearity test: Consistency and asymptotic normality of the auxiliary regression

As explained in Section 2.3.2, the linearity test is based on the auxiliary regression (2.4) which for observation i can be written as

$$Y_i = \mu_i \iota + X_i^* \beta + R_i \tag{A.1}$$

where $Y_i = (y_{i1}, \ldots, y_{iT})'$, X_i^* is a $(T \times km)$ matrix with $X_i^* = (X_i, (X_i \odot q_{it}), \ldots, (X_i \odot q_i^m))$ and R_i is the error vector. After eliminating the fixed effects, the least squares estimator of β has the form

$$\begin{aligned} \hat{\beta} &= (\sum_{i=1}^{N} X_i^{*\prime} Q_T X_i^*)^{-1} \sum_{i=1}^{N} X_i^{*\prime} Q_T Y_i \\ &= \beta^0 + (\sum_{i=1}^{N} X_i^{*\prime} Q_T X_i^*)^{-1} \sum_{i=1}^{N} X_i^{*\prime} Q_T R_i \end{aligned} \tag{A.2}$$

Under the null hypothesis, $R_i = U_i$ for $i = 1, \ldots, N$. Moreover, some elements of β^0 equal zero under H_0, but the idea behind the linearity test is that these elements can be estimated consistently. Thus, under H_0,

$$\begin{aligned}\hat{\beta} &= \beta^0 + (\sum_{i=1}^{N} X_i^{*\prime} Q_T X_i^*)^{-1} \sum_{i=1}^{N} X_i^{*\prime} Q_T U_i \\ &= \beta^0 + \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} c_{ts} x_{it}^* x_{is}^{*\prime} \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} c_{ts} x_{it}^* u_{is} \end{aligned} \tag{A.3}$$

The second equality follows from Lemma 1. Formula (A.3) provides a convenient form for analyzing the asymptotic properties of $\hat{\beta}$ when $N \to \infty$ and T is fixed. In fact, consistency and asymptotic normality can be established in the usual way by applying the Markov law of large numbers and the Lyapunov central limit theorem; See Theorems 3.15 and 5.13 in White (2000).

A.2 Consistency and asymptotic normality of the maximum likelihood estimator in the Panel Smooth Transition model

In this appendix we sketch the proof of Theorem 2 in the chapter.

A.2.1 Logarithmic likelihood, score and Hessian

As described in Section 3.3 the estimation of PSTR models is carried out in two steps. First, the fixed effects are eliminated by multiplying (2.1') with the within transformation matrix Q_T. In the second step, the log-likelihood, after eliminating the fixed effects, is maximized with respect to $\beta = (\beta_1', \beta_2', \beta_3')'$. The concentrated log-likelihood can be written as follows:

$$L_N(\beta) = c - \frac{1}{2N} \sum_{i}^{N} (Y_i - X_i\beta_1 - W_i(\beta_3)\beta_2)' Q_T (Y_i - X_i\beta_1 - W_i(\beta_3)\beta_2). \tag{A.4}$$

The elements in score vector $S_N(\beta)$ for (A.4), evaluated at the true value

$\beta = \beta^0$, are,

$$\frac{\partial L(\beta)}{\partial \beta_1}\Big|_{\beta=\beta^0} = \frac{1}{N}\sum_{i=1}^{N} X_i' Q_T U_i$$

$$\frac{\partial L(\beta)}{\partial \beta_2}\Big|_{\beta=\beta^0} = \frac{1}{N}\sum_{i=1}^{N} W(\beta_3^0)' Q_T U_i$$

$$\frac{\partial L(\beta)}{\partial \beta_{3j}}\Big|_{\beta=\beta^0} = \frac{1}{N}\sum_{i=1}^{N} \beta_2^{0\prime}\frac{\partial W_i(\beta_3^0)'}{\partial \beta_{3j}} Q_T U_i \quad j = 1, \ldots, m$$

The elements of the Hessian matrix have the form

$$\frac{\partial^2 L(\beta)}{\partial \beta_1 \beta_1'} = -\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T X_i \tag{A.5}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_1 \beta_2'} = -\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T W_i(\beta_3) \tag{A.6}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_1 \beta_{3j}} = -\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T \frac{\partial W_i(\beta_3)}{\partial \beta_{3j}}; \quad j = 1, \ldots, m \tag{A.7}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_2 \beta_2'} = -\frac{1}{N}\sum_{i=1}^{N} W_i(\beta_3)' Q_T W_i(\beta_3) \tag{A.8}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_2 \beta_{3j}} = -\frac{1}{N}\sum_{i=1}^{N} \left[\beta_2' \frac{\partial W_i(\beta_3)'}{\partial \beta_{3j}} Q_T W_i(\beta_3) - \beta_2' \frac{\partial W_i(\beta_3)'}{\partial \beta_{3j}} Q_T U_i\right], j = 1, \ldots, m \tag{A.9}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_{3j} \beta_{3h}} = -\frac{1}{N}\sum_{i=1}^{N} \left[\beta_2' \frac{\partial W_i(\beta_3)'}{\partial \beta_{3j}} Q_T \frac{\partial W_i(\beta_3)}{\partial \beta_{3j}} \beta_2 - \beta_2' \frac{\partial^2 W_i(\beta_3)'}{\beta_{3j}\beta_{3h}} Q_T U_i\right], \tag{A.10}$$

$j, h = 1, \ldots, m.$

A.2.2 Consistency

In order to proof consistency of the maximum likelihood estimator $\hat{\beta}$ we apply Theorem 4.1.1 in Amemiya (1985). The maximum likelihood estimator is consistent if (i) the parameter space Ω is compact, (ii) the objective function is continuous in $\beta \in \Omega$ and (iii) the objective function converges to a non-stochastic function uniformly in probability in $\beta \in \Omega$ as $N \to \infty$. Conditions

(i) and (ii) are satisfied given assumptions (E1) to (E7). In what follows we verify that (*iii*) is also satisfied.

Using Assumption (E1) and the equality

$$W_i(\beta_3^0)\beta_2^0 - W_i(\beta_3)\beta_2 = -W_i(\beta_3)(\beta_2 - \beta_2^0) - (W_i(\beta_3) - W_i(\beta_3^0)\beta_2^0$$

one can write the likelihood function (A.4) in deviations from the mean as,

$$\begin{aligned}
L_N =& \frac{1}{N}\sum_{i=1}^{N}(\beta_1 - \beta_1^0)' X_i' Q_T X_i (\beta_1 - \beta_1^0) \\
&- \frac{2}{N}\sum_{i=1}^{N}(\beta_1 - \beta_1^0)' X_i' Q_T W_i(\beta_3)(\beta_2 - \beta_2^0) \\
&- \frac{2}{N}\sum_{i=1}^{N}(\beta_1 - \beta_1^0)' X_i' Q_T (W_i(\beta_3) - W_i(\beta_3^0))\beta_2 \\
&- \frac{2}{N}\sum_{i=1}^{N}(\beta_1 - \beta_1^0)' X_i' Q_T U_i \\
&+ \frac{1}{N}\sum_{i=1}^{N}(\beta_2 - \beta_2^0)' W_i(\beta_3)' Q_T W_i(\beta_3)(\beta_2 - \beta_2^0) \\
&- \frac{2}{N}\sum_{i=1}^{N}(\beta_2 - \beta_2^0)' W_i(\beta_3)' Q_T (W_i(\beta_3) - W_i(\beta_3^0))\beta_2 \\
&- \frac{2}{N}\sum_{i=1}^{N}(\beta_2 - \beta_2^0)' W_i(\beta_3)' Q_T U_i \\
&+ \frac{1}{N}\sum_{i=1}^{N}\beta_2' (W_i(\beta_3) - W_i(\beta_3^0))' Q_T (W_i(\beta_3) - W_i(\beta_3^0)) \\
&- \frac{2}{N}\sum_{i=1}^{N}\beta_2' (W_i(\beta_3) - W_i(\beta_3^0))' Q_T U_i \\
&+ \frac{1}{N}\sum_{i=1}^{N} U_i' Q_T U_i
\end{aligned}$$

In a compact parameter space uniform convergence of the likelihood func-

tion follows from uniform convergence of the following moment matrices:

$$\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T X_i \rightarrow \mathsf{E}\left[X_i' Q_T X_i\right] \tag{A.11}$$

$$\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T W_i(\beta_3) \rightarrow \mathsf{E}\left[X_i' Q_T W_i(\beta_3)\right] \tag{A.12}$$

$$\frac{1}{N}\sum_{i=1}^{N} W_i(\beta_3)' Q_T W_i(\beta_3) \rightarrow \mathsf{E}\left[W_i(\beta_3)' Q_T W_i(\beta_3)\right]. \tag{A.13}$$

Convergence in probability as $N \rightarrow \infty$ of (A.11) follows from Assumption (E6) and the law of large numbers for i.i.d random variables. In order to proof convergence in probability of the moment matrices (A.12) and (A.13) we use the uniform law of large number for i.i.d processes (Theorem 4.5.2 in Amemiya). We have to show that there exists a dominant function $h(X_1, X_2)$ such that the absolute value of the elements of the moment matrices (A.12) and (A.13) are less than $h(X_1, X_2)$ for all $\beta \in \Omega$ and that $\mathsf{E}\left[h(X_1, X_2)\right] < \infty$.

In order to find the dominant function it is convenient to write explicitly the typical (j, h)-element of (A.12) and (A.13). Using Lemma 1, it is seen that the (j,h)-element, $j, h = 1, \ldots, k$ in (A.12) can be written as,

$$\frac{1}{N}\sum_{i=1}^{N} z_i^{(j,h)}(\beta_3) \tag{A.14}$$

where

$$z_i^{(j,h)}(\beta_3) = \sum_{t=1}^{T}\sum_{s=1}^{T} c_{st}\, g(x_{2t}'\beta_3)\, x_{is,h}\, x_{it,j}. \tag{A.15}$$

One dominant function $h(X_1, X_2)$ for $z_i^{(j,h)}(\beta_3)$ is $\sup_{\beta\in\Omega} |z_i^{(j,h)}(\beta_3)|$. In fact, given that $g(x_{2t}'\beta_3) < \eta$, $\eta \in [0, 1]$. We have that, for any $T < T_0 < \infty$,

$$\begin{aligned}
\mathsf{E}[\sup_{\beta\in\Omega} |z_i(\beta_3)|] &\leq \eta \sum_{s=1}^{T} |c_{st}||x_{is,h}\, x_{it,j}| \\
&\leq \eta \sum_{s=1}^{T}\sum_{t=1}^{T} |c_{st}|\mathsf{E}[|x_{is,h}\, x_{it,j}|] \\
&\leq \eta \sum_{s=1}^{T}\sum_{t=1}^{T} |c_{st}|(\mathsf{E}[|x_{is,h}|^2])^{1/2}(\mathsf{E}[|x_{it,j}|^2])^{1/2} \\
&\leq \eta\Delta \sum_{s=1}^{T}\sum_{t=1}^{T} |c_{st}| < \infty.
\end{aligned}$$

where $\Delta < \infty$. A similar argument can be used to show that the sample moment (A.13) converges uniformly to its population moment.

Consequently, conditions (i),(ii) and (iii) of the Theorem 4.1.1 in Amemiya (1985) are satisfied and the maximum likelihood estimator is consistent for $N \to \infty$ and fixed T.

A.2.3 Asymptotic normality

To prove asymptotic normality we apply Theorem 4.1.3 in Amemiya (1985). It states that if (i) $\hat{\beta}$ is consistent, (ii) the score vector evaluated at the true value of the parameters is asymptotically normal (iii) the Hessian matrix is continuous and (iv) the average Hessian converges in probability to a nonsingular matrix for any estimator $\beta^* \to \beta^0$ then $\sqrt{N}(\hat{\beta} - \beta^0) \xrightarrow{d} \mathsf{N}[0, \sigma^2 V^{-1}]$.

Normality of the score evaluated at β^0 follows from the central limit theorem for i.i.d random variables and assumptions (E1) - (E7). The continuity condition (iii) is satisfied given the specification for the panel smooth transition model. To prove the convergence in probability of the average Hessian we apply Theorem 4.2.1 in Amemiya (1985). That is, convergence of the Hessian follows from uniform convergence of its elements and the fact that $\hat{\beta}$ is consistent.

The proof of uniform convergence of the elements (A.5) -(A.8) of Hessian is the same as the proof of uniform convergence of the likelihood function. In order to prove uniform convergence of (A.9) and (A.10), use Lemma 1 and write (A.9) extensively for each element as,

$$\frac{1}{N}\sum_{i=1}^{N}\left[\beta_2^{*\prime}\frac{\partial W_i(\beta_3^*)'}{\partial \beta_{3j}}Q_T W_i(\beta_3^*) - \beta_2^{*\prime}\frac{\partial W_i(\beta_3^*)'}{\partial \beta_{3j}}Q_T U_i^*\right] = \frac{1}{N}\sum_{i=1}^{N}\left[z_{1i}^{(j,h)}(\beta^*) - z_{2i}^{(j,h)}(\beta^*)\right] \tag{A.16}$$

where

$$z_{1i}^{(j,h)}(\beta^*) = \sum_{t=1}^{T}\sum_{s=1}^{T} c_{st}\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)\beta_{2k}^* x_{2is,j}x_{is,k}x_{it,h} \tag{A.17}$$

$$z_{2i}^{(j,h)}(\beta^*) = \sum_{t=1}^{T}\sum_{s=1}^{T} c_{st}\lambda_{is}(\beta_3^*)\beta_{2k}^* x_{2is,j}x_{is,k}u_{it}^* \tag{A.18}$$

for $j = 1, \ldots, m; h, k = 1, \ldots, K$. The term $\lambda_{it}(\beta_3)$ denotes the first derivative of $g(\upsilon)$ with respect to $\upsilon \equiv x_{2t}'\beta_3$, where $x_{2t} = (1, q_{it}, q_{it}^2, \ldots, q_{it}^m)'$. $\lambda_{it}(\beta_3) = (1 - g(x_{2t}'\beta_3))g(x_{2t}'\beta_3)$.

The uniform convergence of (A.16) requires the existence of a dominant function for (A.17) and (A.18). For (A.17) one such dominant functions is $\sup_{\beta\in\Omega}|z_{1i}(\beta^*)|$ which is well defined because the parameter space is compact and the function $\lambda_{it}(\beta_3)$ takes values in the interval [0,1]. Moreover, $\mathsf{E}[\sup_{\beta\in\Omega}|z_{1i}^{(j,h)}(\beta^*)|]$ is finite. In fact,

$$\begin{aligned}\mathsf{E}[\sup_{\beta\in\Omega}|z_{1i}^{(j,h)}(\beta^*)|] \leq & \sum_{t=1}^{T}\sum_{s=1}^{T}|c_{st}|\mathsf{E}[|\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)||\beta_{2k}^*||x_{2is,j}x_{is,k}x_{it,h}|] \\ \leq & \Delta_1\sum_{t=1}^{T}\sum_{s=1}^{T}|c_{st}|(\mathsf{E}[|x_{2is,j}x_{is,k}x_{it,h}|^2])^{1/2}(\mathsf{E}[|\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)|^2)^{1/2} \\ \leq & \Delta_1\Delta_2\Delta_3\sum_{t=1}^{T}\sum_{s=1}^{T}|c_{st}| < \infty\end{aligned}$$

where $|\beta_{2k}| \leq \Delta_1$ and $\mathsf{E}[|\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)|^2] \leq \Delta_2$ and $\mathsf{E}[|x_{2is,j}x_{is,k}x_{it,h}|^2] \leq \Delta_3$. $\Delta_i, i = 1,2,3$ are finite constants. Uniform convergence of $z_{2i}(\beta^*)$ in (A.16) follows from similar arguments.

The proof of uniform convergence of (A.10) is the same but one needs additional moments and notation. In fact, writing the second partial derivative of $W_i(\beta_3)$ as

$$\frac{\partial^2 W_i(\beta_3)}{\partial\beta_{3j}\beta_{3h}} = [1 - 2g(x_{2t}'\beta_3)]\lambda_{it}(\beta_3)x_{2it,j}x_{2it,h} \tag{A.19}$$

one can see that the dominant function for (A.10) exist because the summands of (A.10) are bounded by the $\sup_{\beta\in\Omega}$. Moreover, the supreme is well defined because, the parameter space is compact and the function $|[1 - 2g(x_{2t}'\beta_3)]\lambda_{it}(\beta_3)|$ only takes values in the interval [0,0.1]. For uniform convergence then one needs that $\mathsf{E}[|x_{2it,j}x_{2ist,h}x_{is,r}x_{it,l}|^2] \leq \Delta < \infty$ for $j,h = 1,\ldots,m$, $r,l = 1,\ldots,k$ and $t,s = 1,\ldots,T$.

Given that conditions (i) to (iv) of the Theorem 4.1.3 in Amemiya (1985) are satisfied then we have that

$$\sqrt{N}(\hat{\beta} - \beta^0) \xrightarrow{d} \mathsf{N}[0, \sigma^2 V^{-1}] \tag{A.20}$$

Appendix B

Derivation of the LM test for parameter constancy in PSTR models

We derived the LM statistic by using the local approximation of the concentrated pseudo-likelihood function. That is, we first eliminated the fixed effect from (2.14) and then test the constancy of the remaining parameters. The concentrated pseudo-log likelihood function for observation i is,

$$L = c - \frac{1}{\sigma^2 NT}\sum_{i=1}^{N} e_i^{*\prime} Q_T e_i^* \tag{B.1}$$

where $e_i^* = (Y_i - \iota\mu - X_i\beta_{11} - X_i(\gamma_1, c_i)\beta_{2i} - Z_i(\gamma_1, c_1)\beta^*)$, $X_i(\gamma_1, c_1) = X_i \odot g(q_i, \gamma_1, c_1)$, $Z_i(\gamma_1, c_1) = [W_i \vdots W_i \odot g(q_i, \gamma_1, c)]$, $W_i = X_i \odot S_i$, $S_i = (1/T, \ldots, T/T)'$ and $X_i = (x_{i1}', \ldots, x_{iT}')'$. As before, Q_T denotes the within transformation matrix. The null hypothesis of parameter constancy is H_0 : $\beta^* = 0$.

The average score evaluated at the null can be written as,

$$\begin{aligned} S_N =& \frac{1}{NT\hat{\sigma}^2}\sum_{i=1}^{N}[\hat{V}_i \vdots Z_i(\hat{\gamma}_1, \hat{c}_1)]' Q_T \hat{e}_i^* \\ =& \frac{1}{NT\hat{\sigma}^2}\sum_{i=1}^{N}[0 \vdots Z_i(\hat{\gamma}_1, \hat{c}_1)' Q_T \hat{e}_i^*]' \end{aligned} \tag{B.2}$$

where $\hat{V}_i = [X_i \vdots X_i(\hat{\gamma}_1, \hat{c}_i) \vdots \frac{\partial X_i(\hat{\gamma}_1, \hat{c}_1)}{\partial \gamma_1}\hat{\beta}_{21} \vdots \frac{\partial X_i(\hat{\gamma}_1, \hat{c}_1)}{\partial c_1}\hat{\beta}_{21}]$.

Using the OP estimator for the covariance matrix the LM test is,

$$LM = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{N} \hat{e}_i' Q_T Z_i(\hat{\gamma}_1, \hat{c}_1) \hat{\Sigma}^{-1} Z_i(\hat{\gamma}_1, \hat{c}_1) Q_T \hat{e}_i \tag{B.3}$$

where

$$\hat{\Sigma} = \frac{1}{NT} \sum_{i=1}^{N} [Z_i(\hat{\gamma}_1, \hat{c}_1)' Q_T Z_i(\hat{\gamma}_1, \hat{c}_1) - Z_i(\hat{\gamma}_1, \hat{c}_1)' Q_T \hat{V}_i (V_i' Q_T V_i)^{-1} V_i' Q_T Z_i(\hat{\gamma}_1, \hat{c}_1)] \tag{B.4}$$

Appendix C

Models used in the simulation study

This Appendix contains the different data generating processes used in the Monte Carlo experiments presented in Section 2.4.

The samples for studying the properties of the estimation procedure (Section 2.3.3), the linearity test (Section 2.3.2), the test for no remaining nonlinearity (Section 2.3.4) and the algorithm for determining the number of regimes in the Multiple PSTR model are generated using equation (2.4) while the samples for studying the properties of the parameter constancy test (Section 2.3.4) are generated following equation (2.3.4).

The vector of exogenous variables $(x'_{it}, q_{it})'$ is generated independently for each individual using the following the VAR(1) model,

$$\begin{bmatrix} x_{it} \\ q_{it} \end{bmatrix} = \begin{bmatrix} \Theta & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} x_{it-1} \\ q_{it-1} \end{bmatrix} + \begin{bmatrix} \epsilon^x_{it} \\ \epsilon^q_{it} \end{bmatrix}$$

where x_{it} is a (2×1) vector, Θ is a (2×2) matrix, ρ is a scalar and $(\epsilon^{x\prime}_{it}, \epsilon^q_{it})'$ is a normally distributed vector (3×1) with mean zero and positive definite covariance matrix Σ.

The value of the parameters are $\rho = 0.3$

$$\Theta = \begin{bmatrix} 0.55 & 0 \\ 0 & 0.4 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 0.03 & 0.005 & 0.005 \\ 0.005 & 0.03 & 0.005 \\ 0.005 & 0.005 & 0.3 \end{bmatrix}$$

The value of the parameters for each experiment are presented in Tables C.1 to C.5

Table C.1: Design used for investigating the properties of the estimation method

First experiment ($r = 1, m = 1$)	Second experiment ($r = 1, m = 2$)
$\beta_0 = (1.05, 3.05)'$	$\beta_0 = (1.05, 3.05)'$
$\beta_1 = (2.3, 1.4)'$	$\beta_1 = (2.3, 1.4)'$
$\gamma_1 = 4$	$\gamma_1 = 4$
$c_1 = 3.5$	$c_1 = (3, 4)'$

Table C.2: Design used for investigating the size and power of the linearity test

Size ($r = 1, m = 1$)
$\beta_0 = (1.05, 3.05)'$
$\beta_j = 0$ for $j = 1, \ldots, r$

Power	
First experiment ($r = 1, m = 1$)	Second experiment ($r = 1, m = 2$)
$\beta_0 = (1.05, 3.05)'$	$\beta_0 = (1.05, 3.05)'$
$\beta_1 = (0.021, 0.061)'$	$\beta_1 = (0.021, 0.061)'$
$\gamma_1 = 2$ and 5	$\gamma_1 = 2$ and 5
$c_1 = 3.5$	$c_1 = (3, 4)'$

Table C.3: Design used for investigating the size and power of the test of parameter constancy

Size
$(h = 1)$
$\beta_1 = (1.05, 3.05)'$
$\beta_2 = (0.021, 0.061)'$
$\gamma_1 = 2$
$c_1 = 0.5$

Power	
First experiment $(h = 1)$	Second experiment $(h = 2)$
$\beta_1 = (1.05, 3.05)'$	$\beta_0 = (1.05, 3.05)'$
$\beta_2 = (0.021, 0.061)'$	$\beta_1 = (0.021, 0.061)'$
$\beta_3 = (0.0105, 0.0305)'$	$\beta_3 = (0.0315, 0.0915)'$
$\beta_4 = (0.023, 0.014)'$	$\beta_4 = (0.069, 0.042)'$
$\gamma_1 = 2$	$\gamma_1 = 2$
$c_1 = 0.5$	$c_1 = (0.6, 0.7)'$

Table C.4: Design used for investigating the size and power of the test of no remaining nonlinearity

Size	Power
$r = 1, m = 1$	$r = 1, m = 1$
$\beta_1 = (1.05, 3.05)'$	$\beta_1 = (1.05, 3.05)'$
$\beta_2 = (2.3, 1.4)'$	$\beta_2 = (2.3, 1.4)'$
$\gamma_1 = 3$	$\beta_3 = (0.0315, 0.0915)'$
$c_1 = 3.5$	$\beta_4 = (0.069, 0.042)'$
	$\gamma_1 = 2$ and 5
	$c_2 = 2$

Table C.5: Design used for investigating the properties of the algorithm for determining the number of regimes in a Multiple regime PSTR model

First Experiment ($r = 2, m = 1$)	Second Experiment ($r = 3, m = 1$)
$\beta_1 = (3.5, 3.5)'$	$\beta_1 = (3.5, 3.5)'$
$\beta_2 = (1.9, 1.9)'$	$\beta_2 = (1.9, 1.9)'$
$\beta_3 = (2.5, 3.5)'$	$\beta_3 = (2.5, 3.5)'$
$\gamma_1 = 5$, $c_1 = 2.869$	$\beta_4 = (3.5, 3.5)'$
$\gamma_2 = 10$, $c_2 = 4.354$	$\gamma_1 = 2$, $c_1 = 2.869$
	$\gamma_2 = 9$, $c_2 = 3.479$
	$\gamma_3 = 5$, $c_3 = 4.354$

Chapter 3

Simulation-based finite-sample linearity test against smooth transition models

3.1 Introduction

Monte Carlo (MC) testing technique for testing statistical hypothesis was introduced by Dwass (1957) in statistics and by Dufour (1995) and Dufour and Kiviet (1996) in econometrics. It has been successfully applied in various testing situations in econometrics. For instance, Dufour and Kiviet (1996) apply it to test for structural change in first-order dynamic model, Dufour, Khalaf, and Bernard (*in press*) use the same technique in testing for heteroskedasticity in linear regression models, and Khalaf, Saphores, and Bilodeau (2003) test the null hypothesis of no jumps in a jump process with GARCH errors. Overall, it has been found that the MC testing technique is not only a solution to intractable analytical problems but that it also improves the small sample behaviour of existing asymptotic tests.

In this chapter the MC testing technique is applied to testing linearity against smooth transition regression. Perhaps the most commonly used test for this purpose is the LM-type test that Luukkonen, Saikkonen, and Teräsvirta (1988) introduced. This test is not applicable if the number of regressors in the model is large and the number of observations at the same time very small. To give an example, this may be the situation in panels when the time series are short and the number of individuals small. Indeed, the test developed and discussed in this work is suitable for testing linearity within the panel smooth transition regression model presented in Chapter 2. Testing the random walk hypothesis against the smooth permanent surge model as in Chapter 4, constitutes another application.

Investigating the small-sample size and power properties of the test and comparing them with the corresponding properties of the LM-type test of Luukkonen, Saikkonen, and Teräsvirta (1988) forms an essential part of this study. The results indicate that the MC-based test is practically free from size distortion in small samples and has power against the smooth transition alternative.

The chapter is divided in five sections. The next section describe the different linearity tests considered in the chapter. Section 3 explains the Monte Carlo testing techniques and also contains a discussion of computational details. The results of the Monte Carlo experiments can be found in section 4. Section 5 concludes.

3.2 Linearity tests

The linearity tests that we present in the chapter are tests derived against the Logistic Smooth Transition regression [LSTR] model

$$y_t = x'_{1t}\beta_1 + x'_{2t}\beta_2 G(x_{3t};\gamma,c) + \epsilon_t \tag{3.1}$$

where $G(.)$ is the logistic function,

$$G(x_{3t};\gamma,c) = (1+\exp(-\gamma(x_{3t}-c)))^{-1}, \quad \gamma > 0. \tag{3.2}$$

When $x_{1t} \equiv x_{2t}$ the Logistic Smooth Transition regression model (3.1) may be viewed as a linear model whose parameters change smoothly from β_1 to $\beta_1 + \beta_2$ as a function of an exogenous variable x_{3t}. The smoothness of the change is governed by parameter γ in the transition function (3.2). The larger the value of γ the *steeper* the transition is. In fact, when $\gamma \to \infty$, $G()$ becomes a step function and model (3.1) with (3.2) is a switching regression model with two regimes. Parameter c is the location parameter and indicates the centre of the transition. See Teräsvirta (1998) and van Dijk, Teräsvirta, and Franses (2002) for further details on the model.

Linearity hypothesis of (3.1) may be formulated in two ways. Either H_0 : $\gamma = 0$ which means that under the null, the transition function $G()$ becomes constant and the LSTR model is thus linear. This testing problem is non-standard because under the null hypothesis $(\beta'_2, c)'$ are unidentified nuisance parameters. Luukkonen, Saikkonen, and Teräsvirta (1988) derived a linearity test based on Taylor expansions that circumvent this identification problem. The second option is to choose $\mathrm{H}_0 : \beta_2 = 0$ as the null hypothesis, in which case (γ, c) are not identified when the null hypothesis is valid.

3.2.1 Taylor expansion-based linearity test

Luukkonen, Saikkonen, and Teräsvirta (1988) test linearity in the Logistic Smooth transition model (3.1) as $\mathrm{H}_0 : \gamma = 0$. Assuming $x_{1t} \equiv x_{2t}$ in (3.1), the authors circumvent the identification problem by replacing the transition function (3.2) with its third order Taylor expansion evaluated at $\gamma = 0$ and test the equivalent hypothesis $\mathrm{H}_0 : \tilde{\beta}_2 = \tilde{\beta}_3 = \tilde{\beta}_4 = 0$ in the following auxiliary regression

$$y_t = x'_{1t}\tilde{\beta}_1 + (x_{1t}x_{3t})'\tilde{\beta}_2 + (x_{1t}x_{3t}^2)'\tilde{\beta}_3 + (x_{1t}x_{3t}^3)'\tilde{\beta}_4 + e_t^* \tag{3.3}$$

Assuming the existence of the relevant moments, the standard asymptotic theory is applicable because under the null $e_t^* = \epsilon_t$. Moreover, $\hat{\beta} =$

$(\tilde{\beta}_1', \tilde{\beta}_2', \tilde{\beta}_3', \tilde{\beta}_4')'$ can be estimated consistently and the maximum likelihood estimator is asymptotically normal. The LM type statistic for testing linearity can be written as,

$$LM = \frac{1}{\tilde{\sigma}^2}\hat{u}'\tilde{X}_2(\tilde{X}_2'M_{X_1}\tilde{X}_2)^{-1}\tilde{X}_2'\hat{u} \tag{3.4}$$

where $\tilde{X}_2 = (\tilde{x}_{21}', \ldots, \tilde{x}_{2T}')'$, with $\tilde{x}_{2t} = (x_{1t}'x_{3t}, x_{1t}'x_{3t}^2, x_{1t}'x_{3t}^3)'$, $X_1 = (x_{11}', \ldots, x_{1T}')'$ and $M_{X_1} = I - X_1(X_1'X_1)^{-1}X_1'$. $\hat{u}$ is the vector of residuals under the null hypothesis, $\hat{u} = M_{X_1}Y$ with $Y = (y_1, \ldots, y_T)'$.

The LM type statistic (3.4) can be seen as a rather general nonlinearity test just because the same auxiliary regression (3.3) can be obtained from many additive nonlinear models. In fact, in deriving the test Luukkonen, Saikkonen, and Teräsvirta (1988) assume a rather general functional form for the alternative.

Statistic (3.4) has been found to have good size and power in many relevant testing situations. The main drawback of the test is that its size may be distorted when the number of parameters under the alternative is large relative to the sample size. Besides, the test based on the third-order Taylor expansion is not even available for sample sizes smaller than $k_1 + 3k_2$.

3.2.2 Davies type test

In this section, we follow Davies (1977, 1987) and Hansen (1996, 2000) and derive the LM test for $H_0 : \beta_2 = 0$ in (3.1) assuming that $(\gamma, c)'$ are known values. The LM statistic has the form

$$LM(\gamma, c) = \frac{1}{\tilde{\sigma}^2}\hat{u}'X_2(\gamma, c)\left(X_2(\gamma, c)'M_{X_1}X_2(\gamma, c)\right)^{-1}X_2(\gamma, c)'\hat{u} \tag{3.5}$$

where $X_2(\gamma, c) = G(X_3, \gamma, c) \odot X_2$, $X_2 = (x_{21}', \ldots, x_{2T}')'$ and $X_3 = (x_{31}, \ldots, x_{3T})'$. Furthermore, $\hat{u}$ is the vector of residuals under the null hypothesis.

The LM test (3.5) is similar to the test proposed by Hansen (1996, 2000) for testing linearity in threshold autoregressive models. The main difference is that in these models the only unidentified parameter under the null hypothesis is the location parameter, c, usually called the threshold value. Hansen (1996, 2000) derived the asymptotic distribution of the test statistic and proposed a method to draw from it in order to obtain critical values for the test.

Andrews and Ploberger (1994) solve the identification problem by using the SupLM, ExpLM or AveLM test statistics computed over a given parameter space. We also consider the weighed LM test, wLM, which is computed as a

weighted average of the LM test (3.5). The weights are calculated proportional to the magnitude of the LM test: low values of $LM(\gamma, c)$ have lower weight than the high values. This weighting scheme prevents the test from being heavily influenced by redundant values of (γ, c) that may have a negative effect on the power of the other statistics.

3.3 Monte Carlo testing

Now we briefly introduce the Monte Carlo testing technique. For a general treatment and proofs, see Dufour (1995,2002). Let S be any of the tests statistics presented above. These test statistics have critical regions of the form:

$$S \geq c$$

where c is the critical value.

Let Θ_0 be a subset of the parameter space consistent with the null hypothesis. It is said that a test statistic has level α if

$$\sup_{\theta \in \Theta_0} P_\theta[S \geq c] \leq \alpha \tag{3.6}$$

and is of size α if,

$$\sup_{\theta \in \Theta_0} P_\theta[S \geq c] = \alpha$$

In small samples, to solve for c in (3.6) it is necessary to know the finite sample null distribution of S which is typically a complicated function of the observations and usually not available. Monte Carlo testing allows us to construct exact inference in small samples by simulating the distribution of S under the null hypothesis. In fact, when the statistic under the null distribution is free of nuisance parameters it is enough to generate by simulation N i.i.d replications of S under H_0, $S_1, \ldots, S_N$, independently of S_0 and compute the empirical p-value as

$$\mathrm{P}(S_0)_{MC} = \frac{N\hat{G}_N(S_0) + 1}{N + 1} \tag{3.7}$$

where $\hat{G}_N(S_0) = \frac{1}{N}\sum_{j=1}^{N} I_{[0,\infty)}(S_j - S_0)$ and $I_A(z)$ is the indicator function that takes value 1 if $z \in A$ and zero otherwise. If N is chosen so that $\alpha(N+1)$ is an integer, Dwass (1957) and Dufour (1995) have shown that, under H_0,

$$\mathrm{P}[\mathrm{P}(S_0)_{MC} \leq \alpha] = \alpha.$$

When $x_t = (x_{1t}', x_{2t}')'$ in (3.1) contains only exogenous regressors the statistics (3.4) and (3.5) are free of nuisance parameters under H_0. In fact, conditionally on X, the null distribution of (3.4) depends only on the distribution of $w = u/\sigma$ because, $\hat{u} = M_{X_1} u$ and $\hat{\sigma}^2 = u' M_{X_1} u/(T - k_1)$. Thus,

$$\hat{u}/\hat{\sigma} = (T - k_1)^{1/2} \frac{M_{X_1} u}{(u' M_{X_1} u)^{1/2}} = \frac{M_{X_1} w}{(w' M_{X_1} w)^{1/2}}$$

so no knowledge of $\tilde{\beta}_1$ is needed for simulating (3.4) under the null. Similarly, under H_0, (3.5) is independent of $(\beta_1', \gamma, c)'$. Conditionally on X, the null distribution of (3.5) only depends on the distribution of $w = u/\sigma$.

Also note that for given (γ, c) the $LM(\gamma, c)$ statistic (3.5) is equivalent to an LM test for variable addition in the standard linear regression model. It then follows from Proposition 1 and Corollary 2 in Dufour, Khalaf, and Bernard (*in press*) that for any (γ, c), the statistic (3.5) is a scale-invariant function of OLS residuals and as such a pivotal quantity. Since this holds for any value of (γ, c) it is clear that the null distribution of the SupLM, ExpLM, AveLM and wLM tests is also free of nuisance parameters and that it only depends on the data and the distribution of the standardized residuals.

Following Dufour (1995) the small-sample distribution of these statistics can be obtained by simulation as follows:

1. Compute the value of the statistic, S_0, from the original sample.

2. Generate N i.i.d realizations of the LM test (3.4) or (3.5) by replacing $\hat{u}/\hat{\sigma}$ with $\hat{u}_s = M_{X_1} u_s$ where u_s is a draw from the assumed distribution of u/σ. Compute the value of statistic S_j from the simulated sample.

   The number of replications N is typically small but for a given nominal size α, $\alpha(N+1)$ has to be an integer. For example, for $\alpha = 0.05$, $N = 19$ is enough for correcting the size. Greater values of N increase the power of the test.

3. Compute the Monte Carlo p-value (p_{MC}) as

$$p_{MC} = \frac{N\hat{G}_N(x) + 1}{N + 1} \tag{3.8}$$

   where $\hat{G}_N(S_0)$ is defined above.

When $x_t = (x_{1t}', x_{2t}')'$ contains lagged values of the dependent variable, (3.4) and (3.5) are not free of nuisance parameters under the null hypothesis. Dufour (1995) extended the Monte Carlo testing approach to this case. The

method involves maximizing a simulated p-value function over the nuisance parameter space. That is, for each $\theta \in \Theta_0$ generate N i.i.d replications of S, $S_1(\theta), \ldots, S_N(\theta)$ and compute the corresponding p-value,

$$\mathrm{P}_{MC}(S_0|\theta) = \frac{N\hat{G}_N(S_0|\theta) + 1}{N+1}$$

Finally, compute the *maximized Monte Carlo* p-value as $\mathrm{p}_{MMC} = \sup_{\theta\in\Theta_0}[\mathrm{P}_{MC}(S_0|\theta)]$. Dufour (1995), shows that this procedure has level α. This means that $\mathrm{P}[\sup_{\theta\in\Theta_0}[\mathrm{P}_{MC}(S_0|\theta)] \leq \alpha] \leq \alpha$.

Following Dufour (1995,2002) the *maximized Monte Carlo test* can be computed as follows:

1. Choose the restricted subset of the parameter space Θ_0 over which one is to maximize the p-value function. Θ_0 might be the whole parameter space consistent with the null hypothesis such that the model under the null is not explosive. In case of an autoregressive null model, the largest modulus of the roots of the characteristic polynomial is less than unity.

2. Compute S_0 on the observed sample.

3. Generate N i.i.d error vectors $\epsilon = [\epsilon_1, \ldots, \epsilon_T]'$ and standardize them.

4. Compute the maximized p-value

$$\hat{p}_{MMC} = \sup_{\theta\in\Theta_0}[\mathrm{P}_{MC}(S_0|\theta)]$$

In computing $\hat{p}_{MMC}$ it is important to note that for each value of $\theta \in \Theta_0$ one has to compute $\mathrm{P}_{MC}(S_0|\theta)$ by simulating the model under the null. The initial values for the simulated samples are the same as the ones for the original sample. In that sense, one can say that the *maximized Monte Carlo test* to be computed is conditional on the initial observations. Note that S_0 and the generated standardized residuals in step 3 are treated as fixed, which means that the standardized errors are generated only once.

Finally, since $\mathrm{P}_{MC}(S_0|\theta)$ is not a continuous function of θ standard gradient based methods cannot be used to maximize it. Instead, heuristic search methods such as *Simulated Annealing* can be used. [See Goffe, Ferrier, and Rogers (1994) and Brooks and Morgan (1995) for details]. We use the following set-up for the simulating annealing: Start the algorithm at the value of the parameter vector estimated under the null hypothesis and let the initial temperature $T_0 = 2$. The number of function evaluations before temperature

reduction is $N_\epsilon = 200$ and the stopping ϵ is 0.01. This large value of ϵ is due to the fact that the value of N in computing the Monte Carlo p-value is either 19 or 99 which produce p-values with at most three digits of accuracy. The effect of the low initial temperature and this large value ϵ is compensated by the large number of function evaluations before temperature reduction. As a result, for each value of θ the neighbouring parameter space is densely sampled.

3.4 A Monte Carlo experiment

In this section we report results from a Monte Carlo experiment for studying small sample properties of the proposed MC statistic. The section is divided into two subsections. The first subsection presents results of the linearity test when x_t in (3.1) are exogenous regressors. The second one contains size and power results for the LM-type test when $x_t = (x'_{1t}, x'_{2t})'$ in (3.1) contains lags of the dependent variable.

3.4.1 Linearity test with exogenous regressors

When $x_t = (x'_{1t}, x'_{2t})'$ in (3.1) are exogenous, the different LM tests statistics are pivotal under the null hypothesis, and exact inference is therefore available. This opens up the possibility of comparing the power of the exact test based on (3.5) and the one based on (3.3). The latter test should be less powerful than the former one, and the interesting question is how large the power loss will be.

In order to investigate the size properties of the test we generated 5000 independent samples from model (3.1) with $\beta_2 = 0\iota$. We assume that $x_{2t} \equiv x_{1t}$ and generate the $(k_1 \times 1)$ vectors x_{1t} and the scalars x_{3t} from the uniform distribution U(-10,10). The first element in x_{1t} is always unity. The empirical size is computed for the nominal significance level 0.05. To find out the effect the number of regressors on the size of the test, we choose k_1 =2,5 and 9. The P_{MC} is computed using 19 replications from the normal distribution. A larger number of replications could be used, but this small number is enough for obtaining a size-corrected test. We include both the test based on (3.5) and the test based on the auxiliary regression (3.3) (LM-type) in our experiment.

In computing the SupLM it is important to notice that $LM(\gamma, c)$ in (3.5) is a highly erratic function of its arguments (γ, c) which cause problems for conventional optimization procedures. For this reason, we use *Simulated annealing* (SA) discussed in Section 3.3 in the simulation study. The situation

here is different from the one in computing the maximized Monte carlo p-value, and a conservative setup for the SA algorithm will bring good results. An initial temperature of 10^5 together with a large temperature reduction factor (0.85) will normally work. However, it is possible to use a less conservative set-up, if the SA algorithm is started with good initial values for γ and c. They in turn can be found by carrying out a grid search over the (γ, c)-space. Finally, the optimization procedure has to be penalized for negative values of γ and values of c outside the range of the observed transition variable. This last restriction is due to the fact that the covariance matrix of the score used in the LM test (3.5) is not invertible when the transition function only obtains values that are either zero or unity.

Table 3.1 contains the estimated size of the different LM-type tests. The upper panel of the table contains the empirical size of the SupLM, ExpLM, AveLM and wLM test statistics for $T = 25$ and different values of k_1. As can be seen, the estimated size of the test is close to the nominal size and unaffected by the number of regressors.

Table 3.1: Empirical size of the three linearity tests based on 5000 Monte Carlo replications

MC test ($T = 25$)			
Test	$k_1 = 2$	$k_1 = 5$	$k_1 = 9$
ExpLM	0.050	0.050	0.053
AveLM	0.049	0.054	0.051
wLM	0.048	0.048	0.054
SupLM	0.053	0.034	0.041

LM-type				
Asymp	T	$k_1 = 2$	$k_1 = 5$	$k_1 = 9$
	25	0.044	NaN	NaN
	50	0.052	0.050	0.079
	100	0.051	0.057	0.069
MC	T	$k_1 = 2$	$k_1 = 5$	$k_1 = 9$
	25	0.050	NaN	NaN
	50	0.052	0.048	0.048

The number of Monte Carlo replications used to estimate the P_{MC} is 19. NaN indicates that the results are not available because the sample size is too small compared to the number of regressors in (3.3).

The estimated size of the linearity test based on the auxiliary regression (3.3) is shown in the lower panel of Table 3.1. We evaluate the empirical size of the test using both the asymptotic distribution of the test and the exact small sample distribution. Not surprisingly, the estimated size of the LM-type test when the asymptotic distribution is used is close to its nominal size when the number of regressors remains less than or equal to five but becomes distorted when the number of regressors is large. For example, when the auxiliary regression (3.3) contains 36 regressors and $T = 50$, the empirical size of the LM-type test equals 0.079. This outcome differs from the results based on the exact small sample distribution. The LM-type test with the MC procedure always has an empirical size close to the nominal size independently of the number of regressors. For instance, when $k_1 = 9$ and $T = 50$, the empirical size of the LM-type test equals 0.048.

In order to investigate the power properties of the test we generate 5000 replications of model (3.1) with two different values of β_2 and for each β_2 two different values of γ. The elements of β_1 are ones, and β_2 is a vector with all its elements equal to 0.01 and 0.6, respectively. The location parameter c equals the mean of the transition variable x_{3t}. We only consider $k = 5$ and $T = 50$. In order to determine the effect of the number of replications on the power of the MC test we compute the power using two different number of replications, $N = 19$ and 59.

The results can be found in Table 3.2. It can be seen that the ExpLM, AveLM and wLM tests have higher power than the SupLM test, and the differences can be substantial. For example, for $N = 59$, $\gamma = 10$ and $\beta = 0.6\iota$, the power of the SupLM test equals 0.52, whereas it equals 0.82 for the other two. Note that differences in power between the ExpLM, AveLM and wLM tests are negligible. These results thus clearly suggest the use of any one of these three tests instead of the SupLM test. Not only do they have better power than the SupLM test, but they also require far fewer computations and are thus easier to use than the latter.

As already mentioned, the exact MC tests offer an excellent possibility to investigate how much power one loses by approximating the alternative hypothesis by a Taylor expansion instead of testing linearity directly within the relevant STR model. Results shedding light on this question can also be found in Table 3.2. As may be expected, the exact test ("LM-type" in Table 3.2) based on the auxiliary regression (3.3) is less powerful than the corresponding ExpLM, AveLM or wLM tests. Note, however, that it does have about the same power as the SupLM test. The power of the test based on (3.3) is somewhat sensitive to the value of γ, which is not the case for the

Table 3.2: Empirical power of the linearity tests at 5% nominal level.

SupLM				
	$\beta_2 = 0.06$		$\beta_2 = 0.1$	
γ	$N = 19$	$N = 59$	$N = 19$	$N = 59$
2	0.153	0.176	0.462	0.519
10	0.161	0.169	0.473	0.523
ExpLM				
	$\beta_2 = 0.06$		$\beta_2 = 0.1$	
γ	$N = 19$	$N = 59$	$N = 19$	$N = 59$
2	0.297	0.333	0.751	0.803
10	0.302	0.337	0.757	0.811
AveLM				
	$\beta_2 = 0.06$		$\beta_2 = 0.1$	
γ	$N = 19$	$N = 59$	$N = 19$	$N = 59$
2	0.327	0.360	0.769	0.823
10	0.326	0.361	0.767	0.819
wLM				
	$\beta_2 = 0.06$		$\beta_2 = 0.1$	
γ	$N = 19$	$N = 59$	$N = 19$	$N = 59$
2	0.315	0.349	0.775	0.825
10	0.320	0.354	0.779	0.829
LM-type (Monte Carlo Based)				
	$\beta_2 = 0.06$		$\beta_2 = 0.1$	
γ	$N = 19$	$N = 59$	$N = 19$	$N = 59$
2	0.157	0.125	0.464	0.506
10	0.170	0.194	0.457	0.504

The table contains the empirical power of the linearity test at 5% nominal level. The sample size is 50 and $k_1 = k_2 = 5$. N denotes the number of Monte Carlo replications used in computing the P_{MC}.

other tests. These results accord with the ones reported in Hansen (1996). In that paper, the alternative hypothesis was the threshold autoregressive model, and the tests were based on asymptotic distribution theory.

3.4.2 Linearity test in dynamic smooth transition models

In this subsection we present the results of the Monte Carlo experiment for the LM-type test when $x_t = (x_1', x_2')'$ in (3.1) contains only lags of the dependent variable. As already mentioned, in this set-up (3.4) and (3.5) are not free of nuisance parameters under the null hypothesis. In fact, when the null hypothesis holds the parameters of the linear autoregressive model are nuisance parameters. We do not consider the SupLM, ExpLM, AveLM and

wLM tests because of the large number of computations involved in obtaining the *maximized Monte Carlo* p-values. However, this does not mean that these tests are not available.

In order to evaluate the empirical size of the MC tests we generate 5000 samples of size 30 and 50 from AR(p) models. Since the size of the linearity test (3.4) may be affected by both the number of parameters under the alternative and the magnitude of modulus of the roots of the characteristic polynomial, we consider three different AR(p) processes. The first two are AR(1) models in which the autoregressive parameter equals 0.5 or 0.99. The third one is an AR(6) model in which the largest modulus of the root of the characteristic polynomial is 0.98[1]. This is a extreme configuration for the model under the null, because the DGP is close to be non stationary. It is therefore very likely, that in small samples, the data exhibits a non stationary behaviour.

In order to compare the results of the Monte Carlo testing with other frequently used methods we compute the rejection frequencies of the test using four different estimators for the null distribution. These are the χ^2 and F distributions, valid when $T \to \infty$, the Monte Carlo distribution and the empirical distribution obtained by the bootstrap. The bootstrap distribution is estimated in a way similar to the Monte Carlo distribution, the main difference being that instead of maximizing the p-value function over the nuisance parameter space, we set $\tilde{\beta}_1$ in (3.3) to $\hat{\theta}$, where $\hat{\theta}$ is the LS estimator of $\tilde{\beta}_1$ under the null [2].

Table 3.3a: Size of linearity test in dynamic models

Null model AR(1)								
Sample Size	$T = 30$				$T = 50$			
N	$N = 19$		$N = 99$		$N = 19$		$N = 99$	
$\rho =$	0.5	0.99	0.5	0.99	0.5	0.99	0.5	0.99
MMC	0.006	0.016	0.014	0.027	0.011	0.015	0.019	0.027
Boot	0.049	0.057	0.048	0.056	0.050	0.057	0.049	0.060

Table 3.3 contains the rejection frequencies of the test. As can be seen, the rejection frequencies under the null hypothesis of the Monte Carlo test are

[1] $\rho = (3.476, -4.942, 3.68, -1.52, 0.33, -0.032)'$. The modulus of the roots of the characteristic polynomial are (0.98157,0.63977,0.28486).

[2] This is not exactly parametric bootstrapping. The approach is usually called *local Monte Carlo*. It is exact only when the sample size tends to infinity. Consequently, in finite samples there is no guarantee that the test actually has the appropriate size. See Dufour (2002)

Table 3.3b: Size of linearity test in dynamic models

Sample Size	$T = 30$		$T = 50$	
$\rho =$	0.5	0.99	0.5	0.99
χ^2	0.042	0.055	0.032	0.064
F	0.040	0.052	0.033	0.062

Table 3.3c: Size of linearity test in dynamic models

Null model AR(6)				
	$T = 30$		$T = 50$	
χ^2	0.303		0.284	
F	0.479		0.232	
	$N = 19$	$N = 99$	$N = 19$	$N = 99$
MMC	0.002	0.014	0.002	0.014
Boot	0.063	0.053	0.052	0.054

all well below 0.05, which is the nominal level of the test. This shows that the *maximized Monte Carlo p-value* actually achieves effective size control. It is also clear that the size the of the test when the asymptotic distributions are used is affected by the number of regressors under the alternative and the value of the roots of the characteristic polynomial. In fact, for AR(6) with $T = 50$, the empirical size of the test when the χ^2 and the F distribution are used is 0.28 and 0.23, respectively. This result is the combination of two factors: First, the number of regressors under the alternative is large relatively to the sample size and second, the largest root of the characteristic polynomial is 0.98 which is very close to 1. This does not mean that the asymptotic distribution cannot be used. On the contrary, when the number of regressors is small the results are very good as can be seen from the AR(1) model with $\rho = 0.5$, in which the size with the χ^2 distribution is 0.042.

The above results suggest that the following testing strategy is applicable in our case. First, compute the test from the original sample and test the null hypothesis using the asymptotic distribution. If linearity is rejected use the *local Monte Carlo p-value.* If the null is rejected then use the *maximized Monte Carlo* approach. This strategy is computationally attractive and achieves size control for the test. A similar strategy is proposed by Dufour (1995).

To investigate the power of the *maximized Monte Carlo p-value* test-

Table 3.4: Power Monte Carlo test in dynamic models

	$T = 30$	$T = 50$
MMC	0.085	0.155

The table contains the empirical power of the linearity test at 5% nominal level. The data is generated from the following Smooth Transition autoregressive model: $y_t = 0.5y_{t-1} + (1 + exp(10y_{t-1}))^{-1}(-0.7y_{t-1}) + \epsilon_t$, where ϵ_t is distributed as N(0,1)

ing procedure, we generate 5000 samples from $y_t = 0.5y_{t-1} + (1 + exp(10y_{t-1}))^{-1}(-0.7y_{t-1}) + \epsilon_t$, where ϵ_t is distributed as N(0,1). The rejection frequencies are presented in Table 3.4. As can be seen, the power of the MMC test is good. Overall, we can say that the MMC test not only achieves size control but that the power loss associated with this approach is not large.

3.5 Conclusions

In this chapter we apply Monte Carlo testing techniques to test linearity in the Smooth Transition Models. The techniques presented in this chapter can be applied both to static and to dynamic STR models. The methods are computationally intensive and the computations involved can be unfeasible in large samples. For these cases, it would be more appropriate to base the inference on the asymptotic distribution theory.

The simulation results for the case of exogenous regressors encourage practical application: the panel STR model is a good case in point. Additionally, the pivotal property of the LM statistics when the STR model only includes exogenous regressors allows us to investigate the feasibility of the idea of approximating the alternative STR model using a Taylor expansion of the transition function around the null hypothesis. Our simulation results show that the power loss emanating from this approach compared to testing linearity directly against the appropriate STR model is not large. This supports the commonly held view that the test based on the approximate alternative is a useful tool in practice.

In the case of dynamic STR models, the results of the chapter show that the use of asymptotic distribution theory can yield misleading outcomes in small samples. It is also clear that the *maximize Monte Carlo p-value* is a useful testing strategy for small samples and when the asymptotic distribution of the test is not well known.

Bibliography

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

BROOKS, S. P., AND B. MORGAN (1995): "Optimization using simulated annealing," *The Statistician*, 44, 241–257.

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 alternative.," *Biometrika*, 74, 33–43.

DUFOUR, J.-M. (1995): "Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in Econometrics," *Discussion paper, C.R.D.E. Université de Montréal.*

——— (2002): "Monte carlo tests with nuisance parameters with econometric applications," *Journal of Econometrics.*

DUFOUR, J.-M., L. KHALAF, AND J.-T. BERNARD (*in press*): "Simulation-based finite sample test for heteroskedasticity and ARCH effects," *Journal of Econometrics.*

DUFOUR, J.-M., AND J. F. KIVIET (1996): "Exact test for structural change in first-order dynamic models," *Journal of Econometrics*, 70, 39–68.

DWASS, M. (1957): "Modified randomization test for nonparametric hypothesis," *Annals of Mathematical Statistics*, 28, 181–187.

GOFFE, W., G. D. FERRIER, AND J. ROGERS (1994): "Global optimization of statistical functions with simulated annealing," *Journal of Econometrics*, 60, 65–99.

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

——— (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603.

KHALAF, L., J.-D. SAPHORES, AND J.-F. BILODEAU (2003): "Simulation-based exact jump tests in models with conditional heteroskedasticity," *Journal of Economic Dynamics and Control*, 28, 531–553.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

TERÄSVIRTA, T. (1998): "Modelling economic relationships with smooth transition regressions," in *Handbook of applied economic statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. New York: Marcel Dekker.

VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): "Smooth transition autoregressive models - a survey of recent developments," *Econometric Reviews*, 21, 1–47.

Chapter 4

A Smooth Permanent Surge process

4.1 Introduction

Recently the literature of time series analysis has been developing models in which stochastic shocks can have transitory as well as permanent effects. These models close the gap between stationary autoregressive models, in which all shocks are transitory, and models like the random walk in which all shocks are permanent. Example of these models are: the stochastic unit root process by Granger and Swanson (1997), which is an AR(1) process with the autoregressive parameter varying stochastically around one; the autoregressive conditional root by Rahbek and Shephard (2002), in which the autoregressive parameter changes between one and stationarity following a deterministic function of the past observations; the stochastic permanent break model by Engle and Smith (1999) in which the permanence of a given shock is stochastic and depends on its magnitude, and finally, Gonzalo and Martínez (2003) introduce a threshold integrated moving average model in which large shocks are permanent whereas small ones are transitory.

In the present chapter we introduce the Smooth Permanent Surge [SPS] model. The model is a generalization of the stochastic permanent break model by Engle and Smith (1999). The permanent effect of an innovation is stochastic and depends on a deterministic function of past shocks. In the SPS model, small shocks have transitory effects and large shocks may have permanent effects. The model can be seen as an alternative both to the stochastic break model and to the threshold integrated moving average model.

We present three tests in the smooth permanent surge framework. The first is a test for linearity in moving average models. This test follows Brännäs, De Gooijer, and Teräsvirta (1998). The second test is a test of SPS against a random walk and the third is a test against the stochastic permanent break model by Engle and Smith (1999). The performance of these tests in small samples is evaluated by Monte Carlo experiments. Finally, in other to compare our model with the stochastic permanent break model, we apply our method to the same data set and economic problem as the one used in Engle and Smith (1999). That is, we investigate whether stock prices of two companies that belong to the same market move together or not.

The outline of the chapter is the following. In the second section the smooth permanent surge model is introduced and conditions for invertibility of the model are given. The third section describes the proposed tests and explains their implementation. Section 4.4 presents the results of the Monte Carlo investigation. The application to the stock prices is presented in section 4.5. Section 4.6 concludes.

4.2 Smooth Permanent Surge model

The Stochastic Permanent Break [STOPBREAK] process of Engle and Smith (1999) is defined through the following equations:

$$y_t = m_t + \epsilon_t, \quad t = 0, 1, \ldots, T \tag{4.1}$$

where ϵ_t is a stationary martingale difference sequence with respect to $\mathcal{F}_{t-1}$ where $\{\mathcal{F}_t\}$ denotes an increasing σ-algebra adapted to y_t. Furthermore, m_t is a time-varying conditional mean given by

$$m_t = m_{t-1} + q_{t-1}\epsilon_{t-1} \tag{4.2}$$

where q_{t-1} is a function of ϵ bounded by zero and one.

In order to characterize the dynamic properties of the STOPBREAK process it is useful to measure the effect that a given innovation will have on future values of y_t. One such measure is the *permanent effect of an innovation* defined by Engle and Smith (1999) as follows:

$$\lambda_t \stackrel{d}{=} \lim_{k \to \infty} \frac{\partial f(y_t, k)}{\partial \epsilon} \tag{4.3}$$

where $f(y_t, k) \equiv \mathsf{E}(y_{t+k}|\mathcal{F}_t)$, $\epsilon_t = y_t - \mathsf{E}(y_t|\mathcal{F}_{t-1})$. The *permanent effect of an innovation* in the STOPBREAK model is,

$$\begin{aligned} \lambda_t =& q_t + \frac{\partial q}{\partial \epsilon}\Big|_\epsilon \\ =& q_t(1 + \eta_{q,t}) \end{aligned} \tag{4.4}$$

where $\eta_{q,t} \equiv (\partial q/\partial \epsilon\,|_\epsilon)(\epsilon_t/q_t)$.

From (4.4) it is seen that in the STOPBREAK model the long-run effect of an innovation is random and varies over time. The sign and magnitude of the effect depend on the specific functional form of q_t. For instance, if q_t is positive and has positive first derivatives with respect to $|\epsilon|$, $\lambda_t > 0$ for all t, and consequently all shocks have permanent effects.

Further understanding of the role of q_t in STOPBREAK models can be gained by writing (4.1) and (4.2) as an integrated nonlinear moving average model:

$$\Delta y_t = \epsilon_t - \varpi_{t-1}\epsilon_{t-1} \tag{4.5}$$

where $\varpi_{t-1} = 1 - q_{t-1}$. When $q_{t-1} = 1$ for all t, it follows that $\varpi_{t-1} = 0$, which implies that all shocks have permanent effects. On the contrary, when

$q_{t-1} = 0$ for all t, $\varpi_{t-1} = 1$ and all innovations will have a transitory effect on y_t.

Different specifications for ϖ_{t-1} have been proposed in the literature. Engle and Smith (1999) define $\varpi_{t-1} = 1 - q_{t-1}$ with

$$q_{t-1} = \epsilon_{t-1}^2/(\gamma + \epsilon_{t-1}^2). \tag{4.6}$$

In this specification, large positive and negative shocks have large (in absolute value) permanent effects while small shocks have small effects. The main drawback of this specification is that only zero shocks ($\epsilon_{t-1} = 0$) have transitory effects. To eliminate this drawback, Gonzalo and Martínez (2003) proposed the Shock-Exciting Threshold Integrated Moving Average [STIMA] model in which ϖ_{t-1} is an indicator function such that $\varpi_{t-1} = \theta_2$ for $|\epsilon_{t-1}| \leq \kappa$ and $\varpi_{t-1} = \theta_1$ otherwise. Hence, if $\theta_2 \leq 1$ small shocks have only transitory effects. The main disadvantage of this model is that due to the discontinuity of the likelihood function, statistical inference is nonstandard, and conducting inference is computationally expensive. Moreover, the STIMA model implies that shocks of either sign greater that $|\kappa|$ will have large permanent effects.

In this chapter, ϖ_{t-1} is defined as

$$\varpi_{t-1} = \theta_1 + \theta_2 g(\epsilon; \gamma, \mathbf{c}) \tag{4.7}$$

where $g(\epsilon_t, \gamma, \mathbf{c})$ is the following logistic function [see Jansen and Teräsvirta (1996)]:

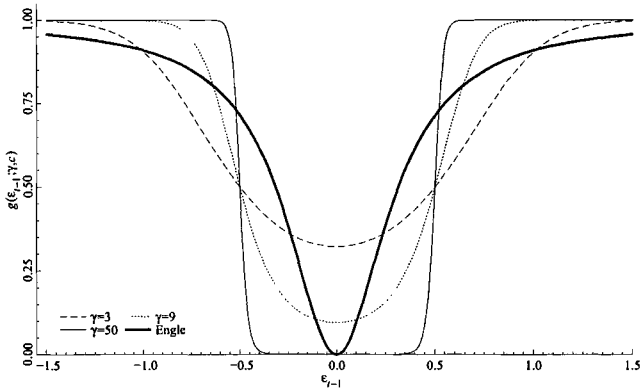
$$g(\epsilon_t; \gamma, \mathbf{c}) = (1 + \exp(-\gamma(\epsilon_t - c_1)(\epsilon_t - c_2)))^{-1} \tag{4.8}$$

with $\gamma > 0$ and $c_1 \leq c_2$. The model defined by (4.5), (4.7) and (4.8) is called Smooth Permanent Surge [SPS] model .

The definition of ϖ_{t-1} in the SPS model has similarities with both the STOPBREAK and the TIMA models. Figure 4.1 plots the transition function (4.8) for different values of γ. For comparison we also include (4.6). As can be seen, the transition function (4.8) has a U-shape form similar to the Logistic function (4.6). However, transition function (4.8) has a broader base. In fact, for relatively large values of γ, the transition function (4.8) practically takes value zero for all $\epsilon_{t-1} > c_1$ and $\epsilon_{t-1} < c_2$, $c_1 < c_2$.

Under some parameter restrictions the behaviour of ϖ_{t-1} in the SPS model approximates the functional form of ϖ_{t-1} in the STOPBREAK one. For instance, when $c_1 = c_2 = 0$, in (4.8), the SPS model approximates the STOPBREAK model quite well. Figure 4.2 shows the transition function (4.8) with $c_1 = c_2 = 0$ together with the function used by Engle and Smith (1999). As can be seen, both transition functions are rather similar for $|\epsilon| > 0$ and attain

Figure 4.1: Transition functions for different γ values



The graph shows different transition functions as a function of ϵ_{t-1}. The thick line is the graph of the transition function (4.6) used by Engle and Smith (1999) with $\gamma = 0.1$. The other lines are the plots of the transition function (4.8) for different value of γ. $c_1 = -0.5$ and $c_2 = 0.5$.

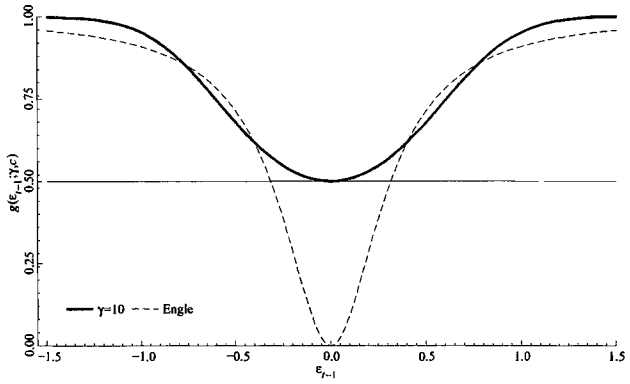
their minimum at $\epsilon = 0$. A difference between them is that the minimum value in the Logistic function (4.6) equals zero whereas it equals $1/2$ in the transition function (4.8). Consequently, if in addition to having $c_1 = c_2 = 0$ in (4.8) we have that $\theta_2 = -1$ and $\theta_2 = 2$ in (4.7) the SPS model is an approximation to the STOPBREAK model.

The SPS model can also approximate the STIMA model. In fact, the STIMA specification can be seen as a limiting case of an unrestricted SPS model. For instance, when $\gamma \to \infty$ and $c_1 = -c_2$ and $c_2 = \kappa > 0$ in (4.8), the SPS model is identical to a STIMA model. This feature of the SPS model makes it possible to describe nonlinearities that the STIMA model also captures. An advantage of the present model over STIMA is that the case $c_1 \neq -c_2$ is included. This feature allows asymmetries between the effects of large positive and negative innovations.

4.2.1 Invertibility

Since the SPS process is a moving average model, the estimation of parameters has to be carried out recursively. In fact, in order to estimate the model one has to be able to estimate the innovation process given the observed data and the generating formula. This is only possible if the model is invertible. Following the definition of invertibility by Granger and Andersen (1978), Engle and

Figure 4.2: Transition function used in the SPS together with the Logistic function used by Engle and Smith (1999)



The graph shows the logistic function used by Engle and Smith (1999) together with the transition function (4.8) when $\gamma = 10, c_1 = c_2 = 0$. The dashed line is the graph of the transition function used by Engle and Smith (1999) when $\gamma = 0.1$. The thick back line represents the transition function (4.8)

Smith (1999) established the invertibility conditions of STOPBREAK models in the following theorem:

Theorem 1 *The nonlinear moving average process in equation (4.5) is invertible if* $E(|1 - q_t(1 + \eta_{qt})|F_{t-1}) \leq z_t < 1$, *where* $\eta_{qt} = (\epsilon_t/q_t)(\partial q/\partial \epsilon)\,|_{\epsilon_t}$ *and* $\{z_t\}$ *is a deterministic sequence defined such that* $\lim_{T\to\infty} \prod_{t=1}^{T} z_t = 0$

Proof. See Engle and Smith (1999). ■

Thus, for invertibility it is required that the average total effect of innovation has to be less than one. Applying Theorem 1 to the SPS model implies that the model is invertible if

$$\mathsf{E}\left[\left|\varpi_{t-1} + \frac{\partial \varpi_{t-1}}{\partial \epsilon_{t-1}}\epsilon_{t-1}\right| \; |\mathcal{F}_{t-1}\right] < 1. \tag{4.9}$$

The invertibility condition has the form,

$$\begin{aligned}\mathsf{E}\left[\left|\varpi_{t-1} + \frac{\partial \varpi_{t-1}}{\partial \epsilon_{t-1}}\epsilon_{t-1}\right| \; |\mathcal{F}_{t-1}\right] &\leq \mathsf{E}_{t-1}|\theta_1 + \theta_2 g(.)| \\ &+\mathsf{E}_{t-1}|\theta_2\gamma(1 - g(.))g(.)\epsilon_{t-1}[(\epsilon_{t-1} - c_1) + (\epsilon_{t-1} - c_2)]|\end{aligned} \tag{4.10}$$

The first term on the right-hand side of (4.10) is less than one if $|\theta_1 + \theta_2| < 1$. The second term is not necessarily zero since its value depends on γ and consequently large values of γ might affect invertibility. Fortunately, when γ is large $g(.)$ tends to a step function taking values zero and one. Hence, for large γ the second term on the right-hand side of (4.10) is practically zero and equals zeros for $\gamma \to \infty$.

4.3 Inference in Smooth Permanent Surge models

This section presents the statistical properties of the SPS model defined by (4.5), (4.7) and (4.8). The random variable of interest is Δy_t and not y_t, which means that the inference is based on the stationary variable Δy_t.

4.3.1 Hypothesis testing

In this subsection we present three tests in the SPS framework. The first test is a test of linearity in (4.5). It is based on Brännäs, De Gooijer, and Teräsvirta (1998). The second test is a test of the random walk hypothesis. The final test is a test of the SPS model against the STOPBREAK alternative.

Testing linearity in the SPS model

Testing linearity in (4.5) is equivalent to testing the hypothesis $\varpi_{t-1} = \theta^*$ for all t. Given that transition function (4.8) is constant when $\gamma = 0$, the linearity test can be carried out by the null hypothesis $H_0 : \gamma = 0$. However, the standard testing procedures are not valid because θ_2, c_1 and c_2 are not identified parameters under the null hypothesis. Brännäs, De Gooijer, and Teräsvirta (1998) circumvent this identification problem following Luukkonen, Saikkonen, and Teräsvirta (1988) and Granger and Teräsvirta (1993). They replace the transition function (4.8) in (4.7) with its first-order Taylor expansion around $\gamma = 0$. After doing that and merging terms it turns out that the null hypothesis $H_0 : \gamma = 0$ in (4.5) is equivalent to $H_0^1 : \tilde{\theta}_2 = \tilde{\theta}_3 = 0$ in the following auxiliary regression:

$$\Delta y_t = \tilde{\theta}_1 \epsilon_{t-1} + \tilde{\theta}_2 \epsilon_{t-1}^2 + \tilde{\theta}_3 \epsilon_{t-1}^3 + e_t^* \tag{4.11}$$

where $e_t^* = \epsilon_t + \epsilon_{t-1} R(\gamma, c; \epsilon_{t-1})$. R is the remainder in the Taylor expansion. Note that under H_0 $e_t^* = \epsilon_t$ so the asymptotic theory is not affected by this approximation.

The LM test statistic is a convenient statistic for testing H_0^1 since it only requires the estimation of a MA(1) process. The resulting LM-type test can be carried out in three steps as follows:

1. Estimate the MA(1) model

$$\Delta y_t = \epsilon_t + \tilde{\theta}_1 \epsilon_{t-1}$$

and compute the residuals $\hat{\epsilon}_t$, $t = 1, \ldots, T$, and the sum of squared residuals SSR_0.

2. Regress $\hat{\epsilon}_t$ on $(\frac{\partial \epsilon_t}{\partial \tilde{\theta}_1}, \frac{\partial \epsilon_t}{\partial \tilde{\theta}_2}, \frac{\partial \epsilon_t}{\partial \tilde{\theta}_3}) \mid_{H_0^1}$ and compute the sum of squares residuals SSR_1. From (4.11) it is seen that the first derivatives of the residuals ϵ_t with respect to $\tilde{\theta}_j$, $j = 1, 2, 3$, under the null hypothesis are of the form

$$\frac{\partial \epsilon_{t-1}}{\partial \tilde{\theta}_j} = \hat{\epsilon}_{t-1} + \hat{\tilde{\theta}}_1 \frac{\partial \epsilon_{t-1}}{\partial \tilde{\theta}_j}$$

and thus have to be computed recursively.

3. The test statistic is,

$$LM = T \frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)}{\mathrm{SSR}_0} \tag{4.12}$$

and has an asymptotic χ^2 distribution with two degrees of freedom under the linearity hypothesis and the assumption $\mathrm{E}\epsilon_t^6 < \infty$.

Testing SPS against random walk

The random walk hypothesis is an interesting one to test, because the behaviour of y_t in the SPS model resembles the behaviour of realizations of the random walk process, and distinguishing between the two is important in applications. In fact, the SPS model can be defined as in (4.5) which is a unit root process with a specific moving average component.

The random walk hypothesis in (4.5) implies $H_0^0 : \varpi_{t-1} = 0$ for all t. This null hypothesis is then equivalent to testing $H_0^1 : \theta_1 = \theta_2 = 0$ in (4.7). The testing problem is again a nonstandard one, because the parameters γ, c are not identified under the null hypothesis. To circumvent the identification problem we follow Davies (1977,1987) and first derive the LM test of $\theta_1 = \theta_2 = 0$ in (4.5) assuming γ and c known. Based in the results in Andrews and Ploberger (1994), the identification problem is solved by applying ExpLM or AveLM tests.

The LM statistic for any given (γ, c) has the form:

$$LM(\gamma, c) = \frac{1}{\hat{\sigma}^2}\hat{u}' X_1(\gamma, c)(X_1(\gamma, c)' X_1(\gamma, c))^{-1} X_1(\gamma, c)'\hat{u} \tag{4.13}$$

where $X_1(\gamma, c) = [\hat{u}_{-1} \vdots G(\hat{u}_{-1}, \gamma, c) \odot \hat{u}_{-1}]$, $\hat{u}$ are the vector of residuals under the null and and $\hat{u}_{-1}$ its first lag, respectively.

The computation of ExpLM and AveLM tests can be based on a dense grid over γ and c. The grid should include possibly large positive values of γ and values of c defined within the range of $\hat{u}$.

In this chapter we do not derive the asymptotic distribution of the test because it is possible to use the small sample distribution. We follow Dufour (1995) and Dufour and Khalaf (2001) and approximate the small sample distribution using Monte Carlo testing techniques. The advantage of this approach is that the test is exact in the sense that it has size-corrected critical regions. The main requirement of the MC test is that the statistic can be simulated under the null hypothesis. Moreover, the test is provably exact when the null distribution is free from nuisance parameters. Statistic (4.13) has this property because when $\mathrm{H}_0 : \theta_1 = \theta_2 = 0$ is valid, we have $\Delta y_t = \epsilon_t$, and consequently all that is needed for simulating Δy_t is the distribution of ϵ_t. Hence, the null distribution of the ExpLM and AveLM tests only depends on the distribution of the errors.

Following Dufour (1995) the small sample distribution of ExpLM and AveLM test can be obtained by simulation as follows:

1. Compute ExpLM or AveLM test from the original sample and call the statistic S_0.

2. Generate the LM test (4.13) by replacing $\hat{u}/\hat{\sigma}$ in (4.13) with $\hat{u}_s = u_s$ where u_s is a draw from the assumed error distribution. Compute the test statistic S_j from the simulated sample. Notice that for simulating the LM-test statistic under H_0, no knowledge of any parameters is needed.

   The number of replications N is typically small but it has to be such that $\alpha(N+1)$ is an integer for a given nominal size α. For example, for $\alpha = 0.05$, $N = 19$ is enough for correcting the size. Greater values of N increase the power of the test.

3. Compute the Monte Carlo p-value (P_{MC}) as

$$\mathrm{P}_{MC} = \frac{N\hat{G}_N(x) + 1}{N+1} \tag{4.14}$$

where $\hat{G}_N(S_0) = \frac{1}{N}\sum_{j=1}^{N} I_{[0,\infty)}(S_j - S_0)$ and $I_A(z) = 1$ for $z \in A$ and 0 otherwise.

The random walk hypothesis can also be tested with the SupLM test. The computation of the SupLM test is difficult because (4.13) is a highly erratic function of (γ, c). Despite this feature of the objective function it can be computed using a suitable global optimization procedure such as *simulated annealing* [See Brooks and Morgan (1995) and Goffe, Ferrier, and Rogers (1994) for details]. The advantage of this algorithm compared to numerical optimization algorithms based on derivatives of the objective function, is that it escapes local optima. The results in Chapter 3 indicate, however, that the ExpLM and AveLM tests have higher power than the SupLM and that they require fewer computations.

Testing STOPBREAK hypothesis within SPS

Even though the STOPBREAK model is not nested in the SPS process, there is a parametrization within the SPS that resembles the characterization of permanent effects in the STOPBREAK process. In the latter model the permanent effect of an innovation λ_t is a random variable defined within [0,2). However, the authors point out the following: *the intuition suggests that the majority of the probability mass for λ_t would lie in the [0,1] interval.* This suggest that even though the STOPBREAK model is invertible for $\lambda_t < 2$ in practice unity serves as upper bound for λ_t.

Using the fact that λ_t is defined on the [0,1] interval and the fact that transition function (4.8) approximates the Logistic function by (4.6) when $c_1 = c_2 = 0$ one can write an approximate STOPBREAK process as follows:

$$\begin{aligned} \Delta y_t &= \epsilon_{t-1} - \varpi_{t-1}\epsilon_{t-1} \\ \varpi_{t-1} &= -1 + 2g(\epsilon_{t-1}, \gamma) \\ g(\epsilon_{t-1}, \gamma) &= \left(1 + \exp(-\gamma\epsilon_{t-1}^2)\right)^{-1}. \end{aligned} \tag{4.15}$$

In (4.15), the *permanent effect of an innovation* λ_t equals

$$\lambda_t = -1 + 2g(\epsilon_{t-1}, \gamma) + 2(1 - g(\epsilon_{t-1}, \gamma))g(\epsilon_{t-1}, \gamma)\gamma\epsilon_{t-1} \tag{4.16}$$

The permanent effect of an innovation in (4.15) lies in the interval [0,1]. In fact, when $\epsilon = 0$, $\lambda_t = 0$ and when $|\epsilon| \to \infty$ we have that $g(.) = 1$ and $\lambda_t = 1$. This means that zero shocks have transitory effects whereas large shocks have permanent effects.

The second term on the right-hand side of (4.16) depends on γ. This might imply that for any ϵ, large values of γ are associated with a large permanent effect. However, when $\gamma \to \infty$, this second term is always zero because the transition function then becomes a step function taking only values zero and one.

This characterization of the STOPBREAK process within the SPS model allows us to formulate the STOPBREAK null hypothesis as $H_0 : c_1 = c_2 = 0$; $\theta_1 = -1$; $\theta_2 = 2$ in (4.5) with (4.7). This hypothesis is not testable because the elements of the score for c_1 and c_2 under the null are the same. However, reformulating the transition function (4.8) as in (4.15) yields

$$g(\epsilon, \gamma, c) = (1 + \exp(-\gamma(\epsilon_{t-1} - c)^2)^{-1}. \tag{4.17}$$

Using this formulation it is possible to test the equivalent null hypothesis $H_0 : c = 0, \theta_1 = -1, \theta_2 = 2$.

Since only γ has to be estimated under the null hypothesis, the LM test is computationally convenient. The test can be obtained in three steps as follows:

1. Estimate (4.5) under the null hypothesis, compute the residuals $\hat{u}$ and the sum of square residuals SSR_0.

2. Estimate the auxiliary regression

$$\hat{u}_t = x_t'\mathrm{b} + \text{error}$$

and compute the sum of squared residuals SSR_1. Where, $x_t = (\frac{\partial \epsilon_t}{\partial \theta_1}, \frac{\partial \epsilon_t}{\partial \theta_2}, \frac{\partial \epsilon_t}{\partial c}, \frac{\partial \epsilon_t}{\partial \gamma})'$.

3. Compute the value of the LM statistic

$$LM = T\frac{(SSR_0 - SSR_1)}{SSR_0}$$

Since the parameter estimates in the SPS model are asymptotically normally distributed, the LM test statistic has an asymptotic χ^2 distribution with three degrees of freedom under the null hypothesis.

The first step in this procedure requires nonlinear estimation of γ. Consequently, the estimated residuals and $\partial\epsilon_t/\partial\gamma$ are not necessarily orthogonal, which can affect the size of test. To circumvent this problem Eitrheim and Teräsvirta (1996) ortogonalized the estimated residuals under H_0 with respect to $\partial\epsilon_t/\partial\gamma$ before computing SSR_1 in step two. The second step in the above algorithm can thus be replaced by the following two steps:

2a. Regress $\hat{u}_t$ on $\frac{\partial \epsilon_t}{\partial \gamma}$ and compute the residuals $\hat{u}_t^o$ and SSR_0.

2b. Estimate the auxiliary regression

$$\hat{u}_t^o = x_t' \text{b} + \text{error}$$

and compute the sum of squared residuals SSR_1.

4.3.2 Estimating SPS models

Invertibility is required for the estimation of the SPS model, and the parameter vector $\varphi = (\theta_1, \theta_2, \gamma, c_1, c_2, \sigma^2)'$ can be estimated by maximum likelihood. The invertibility condition ensures that the likelihood function is well-defined since ϵ_t can be obtained recursively from any initial condition. For practical purposes, we assume that $\epsilon_0 = 0$.

The log likelihood function for an SPS model (4.5) is,

$$L(y, \varphi_1, \sigma^2) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{T}\epsilon_t^2 \tag{4.18}$$

where $\epsilon_t = \Delta y_t + \varpi_{t-1}\epsilon_{t-1}$ and ϖ_{t-1} is defined in (4.7) and (4.8).

The score vector is given by

$$\frac{\partial L}{\partial \beta} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\epsilon_t w_t$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{1}{\sigma^3}\sum_{t=1}^{T}(\epsilon_t^2 - \sigma^2)$$

where $\beta = (\theta_1, \theta_2, \gamma, c)'$ and

$$w_t = b_{t-1}\epsilon_{t-1} + \left(\theta_2 \frac{\partial g_{t-1}(\gamma, \mathbf{c})}{\partial \epsilon}\epsilon_{t-1} + \varpi_{t-1}\right) w_{t-1}$$

and $b_{t-1} = \left(1, g_{t-1}(\gamma, \mathbf{c}), \theta_2 \frac{\partial g_{t-1}(\gamma, \mathbf{c})}{\partial \gamma}, \theta_2 \frac{\partial g_{t-1}(\gamma, \mathbf{c})}{\partial c_1}, \theta_2 \frac{\partial g_{t-1}(\gamma, \mathbf{c})}{\partial c_2}\right)'$.

Furthermore, $g_{t-1}(\gamma, \mathbf{c}) = g(\epsilon_{t-1}; \gamma, \mathbf{c})$ and

$$\frac{\partial g_{t-1}}{\partial \gamma} = [1 - g_{t-1}(\gamma, \mathbf{c})]\, g_{t-1}(\gamma, \mathbf{c})(\epsilon_{t-1} - c_1)(\epsilon_{t-1} - c_2)$$

$$\frac{\partial g_{t-1}}{\partial c_1} = -[1 - g(\epsilon_{t-1})]\, g_{t-1}(\gamma, \mathbf{c})\gamma(\epsilon_{t-1} - c_2)$$

$$\frac{\partial g_{t-1}}{\partial c_2} = -[1 - g_{t-1}(\gamma, \mathbf{c})]\, g_{t-1}(\gamma, \mathbf{c})\gamma(\epsilon_{t-1} - c_1)$$

$$\frac{\partial g_{t-1}}{\partial \epsilon} = [1 - g_{t-1}(\gamma, \mathbf{c})]\, g_{t-1}(\gamma, \mathbf{c})\gamma\left[(\epsilon_{t-1} - c_1) + (\epsilon_{t-1} - c_2)\right].$$

The recursion for computing the likelihood and the score can be started from zero. This starting-value should not have any effect on the results.

The following theorem establishes the consistency and asymptotic normality of the maximum likelihood estimator of φ.

Theorem 2 *Suppose that y_t is generated by an invertible SPS model, where $\{\epsilon_t, \mathcal{F}_t\}$ is a strictly stationary α-mixing martingale sequence. Then the maximum likelihood estimator $\hat{\varphi} \equiv \mathsf{argmax}_{\Psi} L_T(y^T, \varphi)$ is consistent under the following assumptions: (i) $E|\epsilon_t|^{2p} \leq M < \infty$ for some $p > 1$, (ii) the parameter space Ψ is a compact subset of $\mathbb{R}^5$, and (iii) $\varphi_0 = \mathsf{argmax}_{\Psi} E L_T(y^T, \varphi)$ is unique. Moreover, if in addition of (ii) and (iii) the α-mixing coefficients on ϵ_t are of size $p/(1-p)$ and $E|\epsilon_t|^{4p} \leq M < \infty$, then*

$$\sqrt{T}(\hat{\varphi} - \varphi_0) \xrightarrow{d} N(0, V_0^{-1}) \tag{4.19}$$

Proof. The proof of the theorem closely follows Engle and Smith (1999) and can be found in the Appendix A. ■

Even though the parameter estimates are asymptotically normal this result must be applied with caution. In particular, one must be aware of the identification problem involved in testing several null hypotheses. For instance, the standard t-test for θ_2 and γ creates a situation in which the model contains nuisance parameters not identified under the null hypothesis. Consequently, the standard asymptotic distribution of this test is not applicable. For this reason, we recommend that the final estimated model be considered together with the results of the linearity tests.

The second factor that might affect the usefulness of the asymptotic distribution theory is that for very large values of γ the final estimated Hessian may be ill-conditioned. In this situation it is only possible to obtain standard errors for θ_1 and θ_2 by using the corresponding block of the final estimated Hessian.

4.4 Small-sample properties of tests

In this section we investigate the empirical size and power of our tests by simulation. The section is divided in three subsections. The first one is devoted to the linearity test, the second subsection contains results on the small-sample properties of the test of the random walk hypothesis. The final subsection is concerned with the test of the SPS process against the STOPBREAK model. All results are based on 5000 Monte Carlo replications. The data for each

experiment is generated using the following SPS model:

$$\begin{aligned} \Delta y_t &= \epsilon_t + \varpi_{t-1}\epsilon_{t-1} \\ \varpi_{t-1} &= \theta_1 + \theta_2 g(\epsilon_{t-1}, \gamma, c) \\ g(\epsilon_{t-1}, \gamma, c) &= \Big(1 + \exp(-\gamma(\epsilon_{t-1} - c_1)(\epsilon_{t-1} - c_2))\Big)^{-1}. \end{aligned} \tag{4.20}$$

4.4.1 Small-sample properties of the linearity test

In order to investigate the size of the linearity test the data is generated using equation (4.20) with $\theta_2 = \gamma = c_1 = c_2 = 0$. Since only θ_1 is likely to affect the size of the test we consider the test for $\theta_1 = 0.1, 0.2, \ldots, 0.9$. The sample size is 100. The errors ϵ_t are independently normally distributed with mean zero and variance one. The results for the nominal size 0.05 are presented in Table 4.1.

It is seen that the size properties of the test are generally very good in the sense that the empirical size is close to the nominal size. The size deteriorates somewhat for values of θ_1 close to one. A likely explanation is that for θ_1 near one the null model is close to being noninvertible.

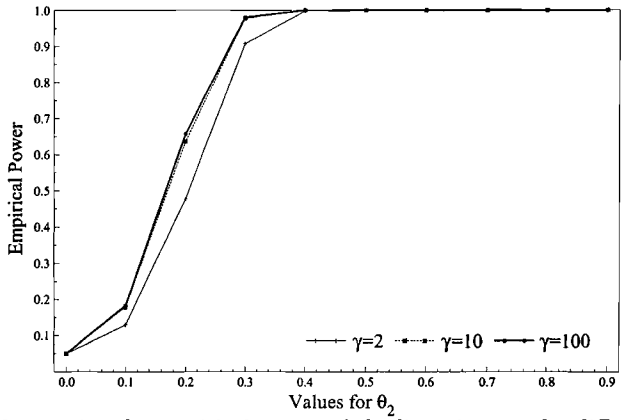
Table 4.1: Empirical size of the linearity test

θ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Size	0.048	0.049	0.049	0.056	0.050	0.054	0.060	0.060	0.085

The table contains the empirical size of the linearity test at 5% nominal level. The sample size is $n = 100$.

In order to investigate the power of the test we generate data from equation (4.20) with $\theta_1 = 0.5$ and different values for θ_2 and γ. We set $\theta_1 = 0.5$ because the empirical size was very close to its nominal size at this value of θ_1. Additionally, we use 10 different values for θ_2 and for each value of θ_2 we compute the power of the test for: γ, $\gamma = 2, 10$ and 100. Figure 4.3 shows the empirical power of the test as a function of θ_2. Each curve in the graph corresponds to one value of γ. As can be seen, the power increases with θ_2. Moreover, for each θ_2 the power is higher for larger values of γ. By comparing the estimated powers for $\gamma = 10$ and $\gamma = 100$ in Figure 4.3, it is seen that both curves are close to each other for all θ_2. This might suggest that the increment in power associated with γ decreases with the value of γ.

Figure 4.3: Empirical power of the linearity test at nominal significance level 0.05 as a function of θ_2 and γ.



The graph contains the empirical power of the linearity test for different values of γ and θ_2. Each line corresponds to one value of γ. The sample size equals 100 and $\theta_1 = 0.5$.

4.4.2 Small sample properties of the test of SPS against random walk

In order to obtain the empirical size of the test we simulate data from (4.20) with $\theta_1 = \theta_2 = 0$. ϵ_t are again drawn from a standard normal distribution. We set the nominal size to equal 0.05 and consequently only use 19 Monte Carlo replications for computing the Monte Carlo p-value. The sample size is 50. In order to compute ExpLM and AveLM test we conduct a grid search on γ and c. The grid for γ includes 100 evenly space values within 2 and 1000 and the grid for c includes 50 different values for c_1 and c_2. Since c_1 and c_2 are exchangeable in the likelihood, we only consider values of c_2 such that $c_2 \geq c_1$. Thus, in computing the ExpLM and AveLM test we evaluate the LM statistic (4.13) $(100 \times (50+1) \times 50)/2$ times for both the original sample and the simulated samples. The result is that the empirical size of both tests, the ExpLM and AveLM, equals 0.0466 which is very close to the nominal size.

For computing the power of the test we simulate data from 18 different variants of model (4.20). In particular, we let θ_1 and θ_2 take values 0.1 and 0.3. Additionally, we also consider two values for γ; $\gamma = 2$ and 10. Since the power of the Monte Carlo tests depends on the number of Monte Carlo samples, N, we use $N = 19$ and $N = 59$ in computing the MC-p-value. The results are summarized in Table 4.2. They indicate that both the ExpLM and

Table 4.2: Empirical power of the test of random walk hypothesis against SPS at the significance level 0.05

N=19									
γ	θ_1	θ_2	ExpLM	AveLM	γ	θ_1	θ_2	ExpLM	AveLM
		0.1	0.126	0.126			0.1	0.128	0.128
2	0.1	0.3	0.301	0.301	10	0.1	0.3	0.322	0.322
		0.5	0.542	0.542			0.5	0.575	0.575
		0.1	0.429	0.429			0.1	0.432	0.432
2	0.3	0.3	0.626	0.626	10	0.3	0.3	0.636	0.636
		0.5	0.774	0.774			0.5	0.786	0.786
		0.1	0.729	0.729			0.1	0.730	0.730
2	0.5	0.3	0.817	0.817	10	0.5	0.3	0.819	0.819
		0.5	0.877	0.877			0.5	0.881	0.881
N=59									
γ	θ_1	θ_2	ExpLM	AveLM	γ	θ_1	θ_2	ExpLM	AveLM
		0.1	0.137	0.137			0.1	0.141	0.141
2	0.1	0.3	0.325	0.325	10	0.1	0.3	0.343	0.343
		0.5	0.597	0.597			0.5	0.627	0.627
		0.1	0.475	0.475			0.1	0.478	0.478
2	0.3	0.3	0.678	0.678	10	0.3	0.3	0.689	0.689
		0.5	0.833	0.833			0.5	0.844	0.844
		0.1	0.789	0.789			0.1	0.788	0.788
2	0.5	0.3	0.873	0.873	10	0.5	0.3	0.875	0.875
		0.5	0.921	0.921			0.5	0.924	0.924

The table contains the empirical power of the test of SPS against random walk. The power is computed for the 5% nominal level. The sample size is $n = 50$. Two different number of Monte Carlo samples were used in computing the Monte Carlo p-values, $N = 19$ and $N = 59$.

the AveLM tests have equal power in small samples. As expected, the power of the test increases with the number of Monte Carlo samples used to compute P_{MC}. The power differences when N is varied are not large. For instance, when $\gamma = 2$, $\theta_1 = 0.1$ and $\theta_2 = 0.3$ the power of the test with $N = 19$ equals 0.30 whereas for $N = 59$ it is 0.32. The power seems to be independent of γ and positively related with θ_1 and θ_2. Finally, considering θ_1 and θ_2 the results indicate that the power of the test depends more on θ_1 than on θ_2. For instance, with $\gamma = 2$, $\theta_2 = 0.1$ an increment of θ_1 from 0.1 to 0.3 increases the power in 2.46%, whereas the same exercise for $\theta_1 = 0.1$ with θ_2 changing from 0.1 to 0.3 only increases the power in 1.37%.

These results indicate that when the null hypothesis is rejected it is useful to test whether θ_2 equals zero. This null hypothesis can be tested using the linearity test proposed in section 4.3.1 or as suggested in Chapter 3.

4.4.3 Small sample properties of the test of SPS against the STOPBREAK model

To compute the size of the test against the STOPBREAK we generate data from (4.20) with $\theta_1 = -1$ and $\theta_2 = 2$, $c_1 = c_2 = 0$. Since the only free parameter under H_0 is γ we use three different values for it: $\gamma = 2, 10$ and 30. We also consider two sample sizes $n = 100$ and $n = 200$. Table 4.3 with four columns contains the results for the Monte Carlo experiment. The first

Table 4.3: Empirical size of the test STOPBREAK against SPS at significance level 0.05

Sample Size	γ	Empirical size	Failures
	2	0.0454	5
100	5	0.0918	238
	30	-	>1000
Sample Size	γ	Empirical size	Failures
	2	0.0528	0
200	5	0.0814	59
	30	-	>1000

and second columns of the table indicate the sample size and the value of γ, respectively. The third column contains the estimated size of the test and the last column indicates the number of discarded draws in the Monte Carlo experiment. These draws are samples for which the test was not available because the covariance matrix of the score was not invertible. The size of the test is good for small values of γ but deteriorates when γ increases. Moreover, the empirical size of the test cannot be computed for $\gamma = 30$. There are two explanations for this outcome. First, when $\gamma \to \infty$ the Hessian becomes noninvertible because the transition function under the null equals one for all ϵ_t different from zero. Second, large values of γ in (4.15) push the model under the null hypothesis towards the boundary of the invertibility condition, because the transition function in (4.9) always equals one.

In summary, the results of the Monte Carlo experiment suggest that the test is only available for small values of γ because the model is not identified under the null hypothesis and because it may not be invertible at large values of γ. The identification problem is not only present in the STOPBREAK approximation to SPS. It is also present in the original version of the STOPBREAK model. In fact, when γ in the (4.6) is close to zero, the logistic function takes value one for practically all $\epsilon_t \neq 0$ and the STOPBREAK model

collapses to a random walk. The results of this Monte Carlo experiment support "second-order logistic" function (4.8) as an alternative parametrization to the simple logistic transition function used by Engle and Smith (1999).

Given the size results we only consider the power of the test for $\gamma = 2$ and we generate data from equation (4.20) with the following other parameter values: $c_1 = -0.1$, $c_2 = 0.1$, $\theta_1 = -0.8$ and $\theta_2 = 1.8$. We estimate the power of the test for two sample sizes $T = 100$ and $T = 200$. The results in Table 4.4. show that the test has good power against the alternative and

Table 4.4: Empirical power of the test of SPS against STOPBREAK at significance level 0.05.

Sample Size	Empirical Power	Sample Size	Empirical Power
100	0.226	200	0.404

The table contains the empirical power of the SPS against STOPBREAK. The power is computed for a 0.05 nominal level

that the power increases with the sample size. However, due to the fact that the size of the test deteriorates with the value of γ and that the test is not likely to be available at large values γ, we recommend caution when using it in applications.

4.5 Application

In this section we illustrate the use of the proposed test statistics and the SPS methodology. The application is borrowed from Engle and Smith (1999). We investigate whether the stock prices of companies that belong to the same market have the tendency to move together. In theory, such prices should move together if they follow the industry behaviour, and they might deviate from each other only temporarily and depending on industry-specific shocks. This theory therefore implies that the ratio of these stock prices should not follow a random walk.

We apply the SPS model to daily price series for Texaco, Mobil, IBM, Microsoft, General Motors, Ford, Coca-Cola and Pepsi. In all cases, stock prices are measured as the closing price of stocks listed in the US market. The observations cover a period from January 1988 to March 2004 and were obtained from Reuters 3000 Xtra.

The random walk test was computed using a grid search over γ and c. We include 100 different values for γ evenly spaced between 2 and 1000 and

25 values for c_1 and c_2 defined between -2.5 and 2.5. The Monte Carlo p-value was computed using 59 samples from a standard normal distribution. In order to compare our results with those of Engle and Smith (1999) we computed the random walk test for two different samples. The first sample comprises observations from January 1988 through December 1995, which is the sample used in Engle and Smith (1999), whereas the second sample contains observations from January 1988 to March 2004. The results are reported in Table 4.5. The table is divided into two panels, one for each sample. Each panel presents the results of the standard LM test and its robustified version, LM_R. The LM_R is robustified against heteroskedasticity by using the HAC estimator of the variance-covariance matrix. We prefer the results of the LM_R because the time series are highly heteroskcedastic. From the first panel of Table 4.5 it can be seen that the random walk hypothesis is not rejected at 5% significance level for the period 1988-1995 for the stock price ratios IBM/MSFT and Texaco/Mobil whereas it is rejected for Cola-Cola/Pepsi and General Motors/Ford. Using the complete sample size the results change. In fact, the random walk hypothesis is only rejected for the price ratio General Motors/Ford.

These results have to be interpreted with caution because the null hypothesis of random walk is the joint hypothesis $\theta_1 = \theta_2 = 0$ in (4.5). Thus it is possible that when the null hypothesis is not rejected the data follows a linear MA(1) process and not necessarily a random walk. It is therefore important to complement the results of the random walk test with those of the linearity test. These results can be found in Table 4.6. As before, the table

Table 4.5: Random Walk test results

	(1988:1-1995:12)		(1988:1-2004:3)	
Ratio	Ordinary	Robust	Ordinary	Robust
IBM/MSFT	1.4(0.383)	0.7(0.733)	7.1(0.017)	1.9(0.167)
GM/FORD	6.7(0.034)	4.5(0.017)	33.1(0.017)	20.9(0.017)
COLA/PEPSI	5.1(0.017)	4.5(0.017)	4.4(0.033)	2.1(0.183)
Texaco/Mobil	6.6(0.017)	1.1(0.500)	6.0(0.017)	2.7(0.120)

Note: The table reports the results based on ExpLM test. P-values in parenthesis. The Monte Carlo p-values were computed using 59 samples from the normal distribution.

is divided into two panels. The left-hand panel contains the results for the period 1988-1995 whereas the results for the complete sample are reported in the right-hand panel. Linearity is generally not rejected, the only exception being the Texaco/Mobil ratio in the sample 1988-1995.

Taking into account the results of both tests, the random walk and the linearity test it is seen that, the stock price ratios Coca-Cola/Pepsi and General Motors/Ford can be characterized by a linear stationary MA(1) while the ratio IBM/MSFT seems to follow a random walk. No evidence in favour of smooth permanent surge is found in any of the stock price ratios. There is no need to fit an SPS model to these price ratio series.

Table 4.6: Linearity test results

	(1988:1-1995:12)		(1988:1-2004:3)	
Ratio	LM	LM_R	LM	LM_R
IBM/MSFT	13.2(0.001)	5.8(0.344)	17.3(0.001)	2.1(0.344)
GM/FORD	1.4(0.484)	1.1(0.581)	0.3(0.838)	0.7(0.716)
COLA/PEPSI	0.1(0.975)	0.1(0.981)	10.0(0.001)	3.6(0.166)
Texaco/Mobil	19.1(0.001)	6.9(0.037)	10.2(0.001)	1.2(0.540)

Note: The column label LM_R contains the results of the robust LM test. The tests are robustified using the standard HAC estimator for the variance.

4.6 Conclusions

In this chapter we have introduced the Smooth Permanent Surge mode which generalizes the STOPBREAK model by Engle and Smith (1999). The new parametrization overcomes a difficulty inherent in STOPBREAK model, namely, that all shocks to the model have permanent effect. In SPS models there is the possibility for shocks to have transitory effects.

The SPS model is also an alternative to STIMA models because, it allows for asymmetries in the long-run effect of the shocks. The continuity of the likelihood function permits the use of standard asymptotic theory when carrying out inference on the model parameters.

We describe three tests of the SPS model that can be used in the modelling process. The first test is a test of non linearity, the second test is a test of SPS against random walk and the final test is a test of SPS against STOPREAK process. The results of the Monte Carlo experiment indicate that the first two tests have good properties in small samples while the last test seems to be of little practical use.

Bibliography

ANDREWS, D. W. K. (1988): "Law of large number for dependent non-identically distributed random variables," *Econometric Theory*, 4, 458–467.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

BRÄNNÄS, K., J. G. DE GOOIJER, AND T. TERÄSVIRTA (1998): "Testing linearity against nonlinear moving average models," *Communications in Statistics - Theory and Methods*, 27, 2025–2035.

BROOKS, S. P., AND B. MORGAN (1995): "Optimization using simulated annealing," *The Statistician*, 44, 241–257.

DAVIDSON, J. (1994): *Stochastic Limit theory.* Oxford University Press, Oxford.

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 alternative.," *Biometrika*, 74, 33–43.

DUFOUR, J.-M. (1995): "Monte carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics," *Discussion paper, C.R.D.E. Université de Montréal.*

DUFOUR, J.-M., AND L. KHALAF (2001): "Monte Carlo test methods in econometrics," in *A companion to theoretical econometrics*, ed. by B. Baltagi, pp. 495–519. Blackwell.

EITRHEIM, Ø., AND T. TERÄSVIRTA (1996): "Testing the adequacy of smooth transition autoregressive models," *Journal of Econometrics*, 74, 59–75.

ENGLE, R. F., AND A. D. SMITH (1999): "Stochastic permanent breaks," *Review of Economics and Statistics*, 81, 553–574.

GOFFE, W., G. D. FERRIER, AND J. ROGERS (1994): "Global optimization of statistical functions with simulated annealing," *Annals of mathematical statistics*, 60, 65–99.

GONZALO, J., AND O. MARTÍNEZ (2003): "Threshold integrated moving average process (Does size matter? Maybe so)," *Preliminary draft.*

GRANGER, C. W. J., AND A. P. ANDERSEN (1978): "On invertibility of time series models," *Stochastic processes and their applications*, 8, 87–92.

GRANGER, C. W. J., AND N. R. SWANSON (1997): "An introduction to stochastic unit-root process," *Journal of Econometrics*, 80, 35–62.

GRANGER, C. W. J., AND T. TERÄSVIRTA (1993): *Modelling nonlinear economic relationships.* Oxford University Press, Oxford.

JANSEN, E. S., AND T. TERÄSVIRTA (1996): "Testing parameter constancy and super exogeneity in econometric equations," *Oxford Bulletin of Economics and Statistics*, 58, 735–763.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

RAHBEK, A., AND N. SHEPHARD (2002): "Inference and ergodicity in the autoregressive root model," *Working paper.*

WOOLDRIDGE, J. M. (1994): "Estimation and inference for dependent processes," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2640–2738, New York. Elsevier Science B.V.

Appendix A

Proof of theorem 2

This appendix contains the proof of Theorem 2 stating consistency and asymptotic normality of the maximum likelihood estimators of the parameters of the SPS model. The log-likelihood function is

$$L_T = \sum_{t=1}^{T} q_t(y^t, \varphi) \tag{A.1}$$

where $q_t(y^t, \varphi) = -\frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\Delta y_t + \varpi_{t-1}\epsilon_{t-1})^2$ and $\varphi = (\theta_1, \theta_2, \gamma, c_1, c_2, \sigma^2)'$. Furthermore, ϖ_{t-1} and $g(\epsilon_{t-1}, \gamma, c_1, c_2)$ are defined in (4.7) and (4.8), respectively.

A.1 Consistency

Repeated substitution of ϵ_{t-1} reveals that $q_t(y_t, \varphi)$ is a function of the increasing sequence $y_i, i = 1, \ldots, t$, and consequently it is a heterogenous sequence of y_t. In order to prove consistency we show that conditions (M.1) - (M.3) of Theorem 4.3 in Wooldridge (1994) are satisfied. To invoke theorem 4.3 in Wooldridge (1994) we need to show that $\{q_t(\epsilon_t, \varphi) : t = 1, \ldots\}$ satisfies the uniform law of large numbers on Ψ and that $q_t(y^t, \varphi)$ is measurable for any $\varphi \in \Psi$. For the uniform law of large numbers to hold we need to verify the following conditions:

1. For each $\tilde{\varphi} \in \varphi$, $\{q_t(y^t, \tilde{\varphi}) : t = 1, 2, \ldots\}$ satisfies the weak law of large numbers.

2. There exists a function $h_t(y^t) \geq 0$ such that

   (a) For all $\varphi_1, \varphi_2 \in \Psi$, $|q_t(y^t, \varphi_1) - q_t(y^t, \varphi_2)| \leq h(y^t) \parallel \varphi_1 - \varphi_2 \parallel$

(b) $\{h(y^t)\}$ satisfies the weak law of large numbers.

The strategy to prove condition 1 is the following. First we will show that for any $\tilde{\varphi} \in \varphi$, $\tilde{\epsilon}$ is L_2-NED, which by Theorem 17.9 in Davidson (1994) implies that $\tilde{\epsilon}^2$ is L_1-NED . We conclude from Theorem 17.5 in Davidson (1994) that $\{\tilde{\epsilon}^2 - \tilde{\sigma}^2\}$ is an L_1-mixingale. Finally, if $\tilde{\epsilon}^2$ is uniformly integrable, it follows from Andrews's (1988) weak law of large numbers that

$$\frac{1}{T}\sum_{t=1}^{T} q_t(\tilde{\epsilon}_t, \tilde{\varphi}) \xrightarrow{p} \mathsf{E}\frac{1}{T}\sum_{t=1}^{T} q_t(\tilde{\epsilon}_t, \tilde{\varphi})$$

for any $\tilde{\varphi} \in \Psi$.

For any $\tilde{\varphi}$ we have that $\tilde{\epsilon}_t = \Delta y_t + \tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}$ where $\tilde{\theta}_{t-1} = \tilde{\theta}_1 + \tilde{\theta}_2 g(\epsilon_{t-1}, \tilde{\gamma}, \tilde{c})$. Recursive substitution of $\tilde{\epsilon}_t$ shows that $\tilde{\epsilon}_t = f_t(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots)$ which is a continuous function of a mixing sequence which is not necessarily mixing. However, $\tilde{\epsilon}$ is L_2-NED. To see this we follow Example 17.4 in Davidson (1994) and approximate f_t with a Taylor expansion about zero with respect to ϵ_{t-j} for $j > m$, where m is a fixed real number. This yields,

$$\tilde{f}_t = \tilde{f}_t^m + \sum_{j=m+1}^{t} \left(\frac{\partial f_t}{\partial \epsilon_{t-j}}\right)^* \epsilon_{t-j}$$

where $*$ denotes the evaluation of the derivatives at points in the interval $[0, \epsilon_{t-j}]$. Note that f_t^m is a measurable approximation of $\mathsf{E}\left[\tilde{f}_t | \mathcal{F}_{t-m}^{t+m}\right]$. Consequently, from Theorem 10.12 in Davidson (1994) we have

$$\left\| \tilde{\epsilon}_t - \mathsf{E}\left[\tilde{f}_t | \mathcal{F}_{t-m}^{t+m}\right] \right\|_2 \leq \|\tilde{\epsilon}_t - f_t^m\|_2 .$$

which implies

$$\left\| \tilde{\epsilon}_t - \mathsf{E}\left[\tilde{f}_t | \mathcal{F}_{t-m}^{t}\right] \right\|_2 \leq \left\| \sum_{j=m+1}^{t} \left(\frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}}\right)^* \epsilon_{t-j} \right\|_2$$

Differentiating $\tilde{f}_t$ w.r.t ϵ_{t-j} yields,

$$\begin{aligned}
\frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}} &= \frac{\partial \Delta y_t}{\partial \epsilon_{t-j}} + \left(\tilde{\theta}_{t-1} + \frac{\partial \tilde{\theta}_{t-1}}{\partial \tilde{\epsilon}_{t-j}}\tilde{\epsilon}_{t-1}\right)\frac{\partial \tilde{\epsilon}_{t-1}}{\partial \epsilon_{t-j}} \\
&= \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j+1}\frac{\partial \Delta y_{t-j+1}}{\partial \epsilon_{t-j}} + \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j}\frac{\partial \Delta y_{t-j}}{\partial \epsilon_{t-j}} \\
&= \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j+1}\left(\varpi_{t-1} + \frac{\partial \theta_{t-j}}{\partial \epsilon_{t-1}}\epsilon_{t-1}\right) + \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j}
\end{aligned}$$

where $\tilde{k}_{t-1} = \left(\tilde{\theta}_{t-1} + \frac{\partial \tilde{\theta}_{t-1}}{\partial \tilde{\epsilon}_{t-1}} \tilde{\epsilon}_{t-1}\right)$. Under invertibility $|\tilde{k}_{t-i}| \leq |\bar{k}| < 1$ and consequently

$$\left| \frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}} \right| \leq 2\bar{k}^j$$

which implies

$$\begin{aligned} \left\| \tilde{\epsilon}_t - \mathsf{E}\left[\tilde{f}_t | \mathcal{F}_{t-m}^t\right] \right\|_2 &\leq \sum_{j=m+1}^{t} \left| \left(\frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}} \right)^* \right| \|\epsilon_{t-j}\|_2 \\ &< 2 \sum_{j=m+1}^{\infty} \bar{k}^j \|\epsilon_{t-j}\|_2 = v_m d_t \end{aligned}$$

where $v_m = 2\sum_{j=t+m}^{\infty} \bar{k}^j$ and $d_t = \|\epsilon\|_2$. Consequently, $\{\tilde{\epsilon}_t\}$ is L_2-NED. It follows from Theorem 17.9 in Davidson (1994) that $\{\tilde{\epsilon}_t^2\} = \{\tilde{\epsilon}_t \tilde{\epsilon}_t\}$ is L_1-NED and from Theorem 17.5 in Davidson (1994), that $\{\tilde{\epsilon}_t^2\}$ is L_1-mixing.

For Andrews's (1988) weak law of large numbers to apply we need $\mathsf{E}|\tilde{\epsilon}_t|^{2p} \leq M < \infty$ for $p > 1$. This result follows from

$$\begin{aligned} \|\tilde{\epsilon}_t\|_{2p} &= \|\Delta y_t + \tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}\|_{2p} \\ &\leq \|\Delta y_t\|_{2p} + \|\tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}\|_{2p} \\ &\leq \|\Delta y_t\|_{2p} + \bar{k}\|\tilde{\epsilon}_{t-1}\|_{2p} \end{aligned}$$

hence $\mathsf{E}\tilde{\epsilon}_t^{2p} \leq \left(\frac{1}{1-\bar{k}}\|\Delta y_t\|_{2p}\right)^{2p}$. Note that $\|\Delta y_t\|_{2p}$ exists since $\|\epsilon_t\| \leq M < \infty$. It follows from Andrews's (1988) weak law of large numbers that

$$\frac{1}{T}\sum_{t=1}^{T} q_t(y^t, \tilde{\varphi}) \to \mathsf{E}\frac{1}{T}\sum_{t=1}^{T} q_t(y^t, \tilde{\varphi})$$

for any $\tilde{\varphi} \in \Psi$.

In other to prove that Condition 2 also holds for SPS models we have to find a dominant function $h(y^t)$ such that,

1. For all $\varphi_1, \varphi_2 \in \Psi$, $|q_t(y^t, \varphi_1) - q_t(y^t, \varphi_2)| \leq h(y^t) \parallel \varphi_1 - \varphi_2 \parallel$.
2. $\{h(y^t)\}$ satisfies the weak law of large numbers.

The mean value approximation to the likelihood function around φ_2 is,

$$q_t(y^t, \varphi) - q_t(y^t, \varphi_2) = \frac{\partial q_t(y^t, \tilde{\varphi})'}{\partial \varphi}(\varphi - \varphi_2)$$

where $\tilde{\varphi}$ indicates that the derivative is evaluated at a point between φ and φ_2. Evaluating the mean value approximation at φ_1 and using the triangle inequality we obtain

$$\left|q_t(y^t,\varphi_1)-q_t(y^t,\varphi_2)\right| \leq \left\langle \frac{\partial q_t(y^t,\tilde{\varphi})}{\partial \varphi'} \right\rangle \|\varphi_1-\varphi_2\|$$

where $\langle x \rangle$ denotes the Euclidean norm of the vector x.

We have that,

$$\begin{aligned}\left\langle \frac{\partial q_t(y^t,\varphi)}{\partial \varphi} \right\rangle &= \Big[\tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{1t}^2+\tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{2t}^2+\tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{3t}^2 \\ &\quad +\tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{4t}^2+\tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{5t}^2+\tilde{\sigma}^{-6}(\tilde{\epsilon}_t^2-\tilde{\sigma}_t^2)^2\Big]^{1/2} \\ &\leq \tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{1t}|+\tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{2t}|+\tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{3t}| \\ &\quad +\tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{4t}|+\tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{5t}|+\tilde{\sigma}^{-3}|(\tilde{\epsilon}_t^2-\tilde{\sigma}_t^2)|\end{aligned} \tag{A.2}$$

where $\tilde{w}_{it}$ for $i=1,\ldots,5$, are the derivatives of $\tilde{\epsilon}_t$ with respect to φ:

$$\begin{aligned}\tilde{w}_{1t} &= \frac{\partial \tilde{\epsilon}_t}{\partial \theta_1}=\tilde{\epsilon}_{t-1}+\tilde{k}_{t-1}\tilde{w}_{1t-1} \\ \tilde{w}_{2t} &= \frac{\partial \tilde{\epsilon}_t}{\partial \theta_2}=\tilde{g}_{t-1}(\tilde{\gamma},\tilde{\mathbf{c}})+\tilde{k}_{t-1}\tilde{w}_{2t-1} \\ \tilde{w}_{3t} &= \frac{\partial \tilde{\epsilon}_t}{\partial \gamma}=\tilde{\theta}_2\frac{\partial \tilde{g}_{t-1}(\tilde{\gamma},\tilde{\mathbf{c}})}{\partial \gamma}\epsilon_{t-1}+\tilde{k}_{t-1}\tilde{w}_{3t-1} \\ \tilde{w}_{4t} &= \frac{\partial \tilde{\epsilon}_t}{\partial c_1}=\tilde{\theta}_2\frac{\partial \tilde{g}_{t-1}(\tilde{\gamma},\tilde{\mathbf{c}})}{\partial c_1}\epsilon_{t-1}+\tilde{k}_{t-1}\tilde{w}_{4t-1} \\ \tilde{w}_{5t} &= \frac{\partial \tilde{\epsilon}_t}{\partial c_2}=\tilde{\theta}_2\frac{\partial \tilde{g}_{t-1}(\tilde{\gamma},\tilde{\mathbf{c}})}{\partial c_2}\epsilon_{t-1}+\tilde{k}_{t-1}\tilde{w}_{5t-1}\end{aligned}$$

where $\tilde{\mathbf{c}}=(c_1,c_2)'$ and $\tilde{g}_{t-1}(.)=g(\tilde{\epsilon}_{t-1},\tilde{\gamma},\tilde{\mathbf{c}})$.

Consider

$$\begin{aligned}|\tilde{\epsilon}_t| &\leq |\Delta y_t|+\bar{k}|\epsilon_{t-1}| \\ &\leq \sum_{i=1}^{t}\bar{k}^{i-1}|\Delta y_{t-i}| \leq (1-\bar{k})\sum_{i-1}^{t}\bar{k}^{i-1}|\epsilon_{t-i}|\end{aligned}$$

and similarly,

$$\begin{aligned}|\tilde{w_{1t}}| &\leq |\tilde{\epsilon}_t|+\bar{k}|\tilde{w}_{1t-1}| \\ &\leq \sum_{i=1}^{t}\bar{k}^{i-1}|\tilde{\epsilon}_{t-i}| \leq \sum_{i=1}^{t}\bar{k}^{i-1}\sum_{j=1}^{t-i}\bar{k}^{j-1}|\epsilon_{t-i-j}|\end{aligned}$$

$$
\begin{aligned}
|\tilde{w}_{2t}| &\leq |\tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}})| + \bar{k}|\tilde{w}_{2t-1}| \\
&\leq \sum_{i=1}^{t} \bar{k}^{i-1} |\tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}})| \leq \sum_{i=1}^{t} \bar{k}^{i-1}
\end{aligned}
$$

$$
\begin{aligned}
|\tilde{w}_{3t}| &\leq \left| \tilde{\theta}_2 \frac{\partial \tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}}}{\partial \gamma} \right| + \bar{k}|\tilde{w}_{3t-1}| \\
&\leq \sum_{i=1}^{t} \bar{k}^{i-1} \left| \tilde{\theta}_2 \frac{\partial \tilde{g}_{t-i}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial \gamma} \right| \\
&\leq K_1 \sum_{i=1}^{t} \bar{k}^{i-1} \left| \frac{\partial \tilde{g}_{t-i}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial \gamma} \right| < K(\tilde{\gamma}, \tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1}
\end{aligned} \tag{A.3}
$$

The last equality in (A.3) holds since $|\frac{\partial \tilde{g}_{t-i}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial \gamma}| < K(\tilde{\gamma}, \tilde{\mathbf{c}}) < \infty$. Similarly, $|\tilde{w}_{4t}| \leq K_2(\tilde{\gamma}, \tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1}$ and $|\tilde{w}_{5t}| \leq K_3(\tilde{\gamma}, \tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1}$ with $K_i(\tilde{\gamma}, \tilde{\mathbf{c}}) < \infty$ for $i = 1, 2, 3$. Thus (A.2) becomes

$$
\begin{aligned}
\langle \frac{\partial q_t(y^t, \varphi)}{\partial \varphi} \rangle \leq \frac{1}{\tilde{\sigma}^2} \Bigg(& \sum_{i=1}^{t} \bar{k}^{i-1} \sum_{j=1}^{t-i} \bar{k}^{i-1} |\epsilon_{t-i-j}| + K_1(\tilde{\gamma}, \tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1} \\
& + K_2(\tilde{\gamma}, \tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1} + K_3(\tilde{\gamma}, \tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1} \Bigg) \Bigg(\sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| \Bigg) \\
& + \frac{1}{\tilde{\sigma}^3} \Big(\sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| \Big)^2 \leq h(y^t)
\end{aligned} \tag{A.4}
$$

where $h(y^t) \equiv \sup_{\varphi \in \varphi} \langle \frac{\partial q_t(y^t, \varphi)}{\partial \varphi} \rangle$. Thus, $h(y^t) \equiv A \left(\sum_{i=1}^{t} \bar{k}^{i-1} \sum_{j=1}^{t-i} \bar{k}^{i-1} |\epsilon_{t-i-j}| + B \right) \times \sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| + \left(\sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| \right)^2$ where A and B are finite constants. It turns out that $h(y^t)$ is a function

of $b_t = \sum_{i=1}^{t} \bar{k}^{i-1}|\epsilon_{t-i}|$ which is L_2-NED in ϵ_t. To see this note that

$$\begin{aligned} \|b_t - \mathsf{E}(b_t|\mathcal{F}_{t-m}^{t+m})\|_2 &= \left\| \sum_{i=m+1}^{t} \bar{k}^{i-1}|\epsilon_{t-i}| \right\|_2 \\ &\leq \sum_{i=m+1}^{t} \bar{k}^{i-1}\|\epsilon_{t-i}\|_2 \\ &< \sum_{i=m+1}^{\infty} \bar{k}^{i-1}\|\epsilon_{t-i}\|_2 = v_m d_t \qquad \text{(A.5)} \end{aligned}$$

(A.6)

From (A.5) it follows that b_t is L_2-NED with constants $d_t = \|\epsilon_t\|_2$ and $v_m = \sum_{i=m+1}^{\infty} \bar{k}^{i-1}$. It follows from Theorem 17.9 in Davidson (1994) that b_t^2 is L_1-NED. Moreover, since $\|\epsilon_t\|_{2p} < \infty$, b_t^2 is uniformly integrable. It follows that $h(y^t)$, which is a function of b_t^2, obeys a uniform weak law of large numbers. Consequently, $q(y^t, \varphi)$ satisfies the conditions of Theorem 4.3 in Wooldridge (1994) and we have that $\varphi \xrightarrow{p} \varphi_0$.

A.2 Asymptotic normality of the maximum likelihood estimator

Using the mean value expansion, the maximum likelihood estimator can be approximated as follows,

$$\begin{aligned} \sqrt{T}(\hat{\varphi} - \varphi^0) &= -\left(\frac{\partial^2 L_T(y^T, \tilde{\varphi})}{\partial\varphi\partial\varphi'}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial q(y^T, \varphi^0)}{\partial\varphi} \\ &= -H_0^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial q(y^T, \varphi^0)}{\partial\varphi} \\ &+ \left[H_0^{-1} - \left(\frac{\partial^2 L_T(y^T, \tilde{\varphi})}{\partial\varphi\partial\varphi'}\right)^{-1}\right] \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial q(y^T, \varphi^0)}{\partial\varphi} \\ &= -H_0^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial q(y^T, \varphi^0)}{\partial\varphi} + o_p(1) \end{aligned}$$

The last equality holds if the sample Hessian evaluated at $\tilde{\varphi}$ converges in probability to the population Hessian. To verify this, we need to show that the hessian obeys the uniform law of large numbers.

Writing the parameter vector as $\varphi = (\varphi_1, \sigma^2)'$ with $\varphi_1 = (\theta_1, \theta_2, \gamma, c_1, c_2)'$ the average Hessian can be written as

$$-\frac{\partial L_T(y^T, \tilde{\varphi})}{\partial \varphi \partial \varphi'} = \begin{bmatrix} \tilde{H}_1 & \tilde{H}_3 \\ \tilde{H}_3' & \tilde{H}_2 \end{bmatrix}$$

with

$$\begin{aligned} \tilde{H}_1 &= \frac{1}{T}\sum_{t=1}^{T} h_{1t} = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\left(\tilde{w}_t\tilde{w}_t' + \tilde{\epsilon}_t\frac{\partial \tilde{w}_t}{\partial \varphi_1}\right) \\ \tilde{H}_3 &= \frac{1}{T}\sum_{t=1}^{T} h_{3t} = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\tilde{w}_t\tilde{\epsilon}_t \\ \tilde{H}_2 &= \frac{1}{T}\sum_{t=1}^{T} h_{2t} = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\left(3\frac{\tilde{\epsilon}_t^2}{\tilde{\sigma}^2} - 1\right) \end{aligned}$$

where $\tilde{w}_t = (\frac{\partial \tilde{\epsilon}_t}{\partial \theta_1}, \frac{\partial \tilde{\epsilon}_t}{\partial \theta_2}, \frac{\partial \tilde{\epsilon}_t}{\partial \gamma}, \frac{\partial \tilde{\epsilon}_t}{\partial c_1}, \frac{\partial \tilde{\epsilon}_t}{\partial c_2})'$.

As in the proof of consistency, we show that $\tilde{H}_1$,$\tilde{H}_2$, $\tilde{H}_3$ obey the uniform law of large numbers. For this purpose, we shall show that the following conditions hold,

1. For each $\tilde{\varphi} \in \Psi$, $\{\tilde{h}_{it} : t = 1, 2, \dots\}$ for $i = 1, 2, 3$, satisfies the weak law of large numbers.

2. For each element $\tilde{h}_{it}^{(j)}$ of $\tilde{H}_i$ for $i = 1, 2, 3$ there exists a function $r(y^t) \geq 0$ such that

   (a) For all φ_1, $\varphi_2 \in \Psi$, $|h_{it}^{(j)}(\varphi_1) - h_{it}^{(j)}(\varphi_2)| \leq r(y^t)\|\varphi_1 - \varphi_2\|$

   (b) $\{r(y^t)\}$ satisfies the weak law of large numbers.

In order to show that condition 1 is satisfied we follow a similar strategy to the one used in the consistency proof. That is, first we show that $\tilde{w}_t$ and $\frac{\partial \tilde{w}_t}{\partial \varphi_1}$ are L_2-NED in ϵ_t. This implies that the summands h_{it} are L_1-NED in ϵ_t. It follows from Theorem 17.5 in Davidson (1994) that they are L_1-mixingales. Finally, the assumption $\|\epsilon_t\|_{4p} \leq M < \infty$ guarantees uniform integrability. Consequently, H_i for $i = 1, 2, 3$, satisfy Andrews's (1988) weak law of large numbers.

In order to show that the summands in H_i for $i = 1, 2, 3$ are L_1-NED in ϵ_t we show that $\tilde{w}_{it}$ for $i = 1, \dots, 5$ are L_2-NED. In particular, we show that,

$$\|\tilde{w}_{1t} - \mathsf{E}[w_{1t}|\mathcal{F}_{t-m}^{t+m}]\|_2 \leq \sum_{j=m+1}^{\infty}\left\|\left(\frac{\partial \tilde{w}_{1t}}{\partial \epsilon_{t-j}}\right)^{*}\epsilon_{t-j}\right\|_2$$

we have that

$$\begin{aligned}\left|\frac{\partial\tilde{w}_{1t}}{\partial\epsilon_{t-j}}\right| &\leq \bar{k}\left|\frac{\partial\tilde{w}_{1t-1}}{\partial\epsilon_{t-j}}\right| + \left(1+\left|\frac{\partial\tilde{k}_{t-1}}{\partial\tilde{\epsilon}_{t-1}}\right||\tilde{w}_{1t-1}|\right)\left|\frac{\partial\tilde{\epsilon}_{t-1}}{\partial\epsilon_{t-j}}\right| \\ &\leq \sum_{i=1}^{j}\bar{k}^{i-1}\left(1+\left|\frac{\partial\tilde{k}_{t-i}}{\partial\tilde{\epsilon}_{t-i}}\right||\tilde{w}_{1t-i}|\right)\left|\frac{\partial\tilde{\epsilon}_{t-i}}{\partial\epsilon_{t-j}}\right| \\ &\leq \bar{k}^{j-1}\sum_{i=1}^{j}\left(1+\left|\frac{\partial\tilde{k}_{t-i}}{\partial\tilde{\epsilon}_{t-i}}\right||\tilde{w}_{1t-i}|\right)\end{aligned}$$

where,

$$\begin{aligned}\left|\frac{\partial\tilde{k}_{t-i}}{\partial\tilde{\epsilon}_{t-i}}\right| &\leq \left|(1+\tilde{\epsilon}_{t-i})\frac{\partial\tilde{\theta}_{t-i}}{\partial\tilde{\epsilon}_{t-i}}+\frac{\partial^2\tilde{\theta}_{t-i}}{\partial^2\tilde{\epsilon}_{t-i}}\right| \\ &\leq \left|(1+\tilde{\epsilon}_{t-i})\tilde{\theta}_2\frac{\partial\tilde{g}_{t-i}}{\partial\tilde{\epsilon}_{t-i}}+\tilde{\theta}_2\frac{\partial^2\tilde{g}_{t-i}}{\partial^2\tilde{\epsilon}_{t-i}}\right| \\ &\leq K_1+K_2|\tilde{\epsilon}_{t-i}| \qquad (A.7)\end{aligned}$$

In (A.7), K_1 and K_2 are finite constants. The last inequality in (A.7) follows from the fact that $|\frac{\partial g_{t-i}}{\partial\epsilon_{t-i}}|$ and $|\frac{\partial^2 g_{t-i}}{\partial^2\epsilon_{t-i}}|$ are bounded functions of $|\epsilon_{t-i}|$. Moreover, from the proof of consistency we have, $|\tilde{w}_t| \leq \sum_{i=1}^{t}\bar{k}^{i-1}|\tilde{\epsilon}_{t-i}|$. Thus,

$$\begin{aligned}\|\tilde{w}_{1t}-\mathsf{E}[\tilde{w}_{1t}|\mathcal{F}_{t-m}^{t+m}]\| &\leq \sum_{j=m+1}^{t}\bar{k}^{j-1}\left[\sum_{i=1}^{j}(K_1+K_2\|\tilde{\epsilon}_{t-i}\|_2)\sum_{i=1}^{t-j}\bar{k}^{i-1}\|\tilde{\epsilon}_{t-j-i}\|_2\right]\|\epsilon_{t-j}\|_2 \\ &< K\sum_{j=m+1}^{\infty}j\bar{k}^{j-1}\|\epsilon_{t-j}\|_2 = \quad v_m d_t\end{aligned}$$

with $v_m = K\sum_{j=m+1}^{\infty}j\bar{k}^{j-1}$ and $d_t = \|\epsilon_t\|_2$. That is, $\tilde{w}_{1t}$ is L_2-NED on ϵ_t. The results for $\tilde{w}_{it}, i = 2,\ldots,5$ follow from similar derivations. Consequently, $\tilde{w}_{it}$ for $i = 1,\ldots,5$ are L_2-NED on ϵ_t.

We shall show that $\partial\tilde{w}_t/\partial\varphi_1$ are L_2-NED. The distinct elements of $\partial\tilde{w}_t/\partial\varphi_1$ are

$$\frac{\partial\tilde{w}_{1t}}{\partial\varphi_i} = \tilde{k}_{t-1}\frac{\partial\tilde{w}_{1t-1}}{\partial\varphi_i}+\tilde{w}_{it-1}+\frac{\partial\tilde{k}_{t-1}}{\partial\varphi_i}\tilde{w}_{1t-1},$$

for $i = 1,\ldots,5$.

$$\frac{\partial\tilde{w}_{2t}}{\partial\varphi_i} = \tilde{k}_{t-1}\frac{\partial\tilde{w}_{2t-1}}{\partial\varphi_i}+\frac{\partial\tilde{g}_{t-1}}{\partial\varphi_i}+\frac{\partial\tilde{k}_{t-1}}{\partial\varphi_i}\tilde{w}_{2t-1},$$

for $i = 2, \ldots, 5$.

$$\frac{\partial \tilde{w}_{ht}}{\partial \varphi_i} = \tilde{k}_{t-1}\frac{\partial \tilde{w}_{ht-1}}{\partial \varphi_i} + \frac{\partial \tilde{k}_{t-1}}{\partial \varphi_i}\tilde{w}_{ht-1} + \tilde{\theta}_2\frac{\partial^2 \tilde{g}_{t-1}}{\partial \gamma \partial \varphi_i}\tilde{\epsilon}_{t-1} + \tilde{\theta}_2\frac{\partial^2 \tilde{g}_{t-1}}{\partial \gamma \partial \varphi_i}\tilde{w}_{it-1},$$

for $i \geq h = 3, 4, 5$, where

$$\frac{\partial \tilde{k}_{t-1}}{\partial \varphi_i} = \frac{\partial \tilde{\theta}_{t-1}}{\partial \varphi_i} + \frac{\partial^2 \tilde{\theta}_{t-1}}{\partial \tilde{\epsilon}_{t-1}\partial \varphi_i}\tilde{\epsilon}_{t-1} + \frac{\partial \tilde{\theta}_{t-1}}{\partial \tilde{\epsilon}_{t-1}}\tilde{w}_{it-1}$$

By proceeding in a manner analogous to that for $w_{it}, i = 1, \ldots, 5$ above, we can show that the elements of $\partial \tilde{w}_t / \partial \varphi_1$ are L_2-NED on ϵ_t. Thus, the summands of H_1 are L_1-NED on ϵ_t. Moreover, since $\tilde{w}_t$, $\partial \tilde{w}_t / \partial \varphi_1$, $\tilde{\epsilon}_t$ have more that two finite moments, we can invoke the weak law of large numbers for L_1-mixingale processes. Thus $\tilde{H}_1 \xrightarrow{p} \mathsf{E}(\tilde{H}_1)$ for all $\tilde{\varphi} \in \Psi$.

Finally, we have to show that there exists a function $r(y^t)$ such that

$$|h_{1t}^{(i,j)}(\varphi_1) - h_{1t}^{(i,j)}(\varphi_2)| \leq r(y^t)\langle \varphi_1 - \varphi_2 \rangle$$

where $h_{1t}^{(i,j)}(\varphi_1)$ denotes the (i,j)-th element of $\tilde{H}_1$ evaluated at φ_1. As before, $r(y^t)$ is such that $\sup_{\varphi \in \Psi} \left\langle \frac{\partial h 1 it(\varphi^*)}{\partial \varphi} \right\rangle \leq r(y^t)$.

We have that

$$\left\langle \frac{\partial \tilde{h}_{1t}^{(i,j)}(\tilde{\varphi})}{\partial \varphi} \right\rangle \leq \sum_{s=1}^{6} \left| \frac{\partial \tilde{h}_{1t}^{(i,j)}(\tilde{\varphi})}{\partial \varphi_s} \right|$$

The elements $|(\partial \tilde{h}_{1t}^{(i,j)} / \partial \varphi_s)|$ are not bounded functions of $\{\tilde{\epsilon}_t\}$, but, $\sup_{\varphi \in \Psi} |\partial \tilde{h}_{1t}^{(i,j)} / \partial \varphi_s|$ is function of the L_1-NED process $\{b_t^2\}$. It follows that $H_1 \xrightarrow{p} \mathsf{E}(H_1)$ uniformly in Ψ.

Consider

$$\tilde{H}_2 = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T} \tilde{w}_t \tilde{\epsilon}_t.$$

From the proof of consistency we have that $\{\tilde{w}_t\}$ and $\tilde{\epsilon}_t$ are L_2-NED on $\{\epsilon_t\}$. From this it follows that $\{\tilde{w}_t \tilde{\epsilon}_t\}$ is L_1-NED on $\{\epsilon_t\}$ and satisfies the weak law of large number for L_1-mixingale processes. It is also possible to show that there exists a dominant function $r(y^t)$ independent of φ that satisfies the weak law of large numbers. That is, $H_2 \xrightarrow{p} \mathsf{E}(H_2)$ uniformly on Ψ. Finally, $H_3 \xrightarrow{p} \mathsf{E}(H_3)$ uniformly on Ψ since $\{\tilde{\epsilon}^2\}$ obeys a weak law of large numbers and $\|\epsilon_t\|_{4p} \leq M < \infty$.

Since $\left(\partial^2 L_T(y^T,\tilde{\varphi})/\partial\varphi\partial\varphi'\right) \xrightarrow{p} \mathsf{E}\left[\partial^2 L_T(y^T,\tilde{\varphi})/\partial\varphi\partial\varphi'\right]$ uniformly on Ψ, and $\left(\partial^2 L_T(y^T,\tilde{\varphi})/\partial\varphi\partial\varphi'\right)$ is a continuous function of φ and $\hat{\varphi} \xrightarrow{p} \varphi^0$ it follows that,

$$\sqrt{T}(\hat{\varphi}-\varphi^0) = -H_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T,\varphi^0)}{\partial\varphi} + o_p(1)$$

where

$$H_0 = -\mathsf{E}\left[\frac{\partial^2 L_t(y^T,\varphi^0)}{\partial\varphi\varphi'}\right] = \begin{bmatrix} \frac{1}{T\sigma_0^2}\sum_{t=1}^{T}\mathsf{E}w_t w_t' & 0 \\ 0 & \frac{2}{\sigma_0^2} \end{bmatrix}$$

Now consider

$$\begin{aligned}\lambda' H_0^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T,\varphi^0)}{\partial\varphi} &= \lambda' H_0^{-1/2}\frac{1}{\sqrt{T}}\begin{bmatrix} -\frac{1}{\sigma_0^2\sqrt{T}}\sum_{t=1}^{T} w_t\epsilon_t \\ \frac{1}{\sigma_0^3\sqrt{T}}\sum_{t=1}^{T}(\epsilon_t^2-\sigma_0^2)\end{bmatrix} \\ &= \lambda' H_0^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\end{aligned}$$

with $\lambda'\lambda = 1$. Using the same argument as Engle and Smith (1999), we have that

$$H_0^{-1/2}T^{-1/2}\frac{\partial L_T(y^T,\varphi)}{\partial\varphi}|_{\varphi=\varphi_0} \xrightarrow{d} N(0,1).$$

EFI
The Economic Research Institute

Reports since 1999
A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title

2004
Andersson, A., Essays in Behavioral Finance.
Grönqvist, E., Selection and Moral Hazard in Health Insurance: Taking Contract Theory to the Data.
Jutterström, M., Att påverka beslut – företag i EUs regelsättande.
Larsson, P., Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.
Lagerwall, B., Empirical Studies of Portfolio Choice and Asset Prices.
Malmsten, H., Properties and Evaluation of Volatility Models.
Nilsson, C., Studies in Environmental Economics: Numerical Analysis of Greenhouse Gas Policies.
Regelexplosionen. Ahrne, G., Brunsson, N. (red).
Salabasis, M., Bayesian Time Series and Panel Models – Unit Roots, Dynamics and Random Effects.
Skallsjö, S., Essays on Term Structure and Monetary Policy.
Söderström, J., Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktorienterade företag.
Talia, K., The Scandinavian Currency Union, 1873-1924 - Studies in Monetary Integration and Disintegration

2003
Andersson, H., Valuation and Hedging of Long-Term Asset-Linked Contracts.
Bergman, M., Essays on Human Capital and Wage Formation.
Damsgaard, N., Deregulation and Regulation of Electricity Markets.
Eklund, B., Four Contributions to Statistical Inference in Econometrics.
Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change. Sundgren, B, Mårtensson, P., Mähring, M., Nilsson, K., (editors)
Globalization and its Enemies, Lundahl, M (editor).
Hakkala, K., Essays on Restructuring and Production Decisions in Multi-Plant Firms.
Holgersson, C., Rekrytering av företagsledare. En studie i hoimosocialitet.
Ivaschenko, I, Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.
Lange, F., Brand Choice in Goal-derived Categories - What are the Determinants?
Le Coq, C., Quantity Choices and Market Power in Electricity Market.
Magnusson, P.R., Customer-Oriented Product Development - Experiments Involving Users in Service Innovation.
Meisiek, S., Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.
Mårtensson, A.. Managing Mission-Critical IT in the Financial Industry.
Nilsson, G., Processorientering och styrning – Regler, mål eller värderingar?

Sandberg, R., Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.
Sturluson, J.T., Topics in the Industrial Organization of Electricity Markets.
Tillberg, U., Ledarskap och samarbete – En jämförande fallstudie i tre skolor.
Waldenström, D., Essays in Historical Finance.
Wallén, U., Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.
Ögren, A., Empirical Studies in Money, Credit and Banking - The Swedish Credit Market in Transition under the Silver and the Gold Standards, 1834 – 1913.

2002
Barinaga, E., Levelling Vagueness – A study of cultural diversity in an international project group.
Berglund, J., De otillräckliga - En studie av personalspecialisternas kamp för erkännande och status.
Bolander, P., Anställningsbilder och rekryteringsbeslut.
Damjanovic, T., Essays in Public Finance.
Ekman, M., Studies in Health Economics – Modelling and Data Analysis of Costs and Survival.
Företagerskan – Om kvinnor och entreprenörskap. Holmquist, C. och Sundin, E (red)
Heyman, F., Empirical Studies on Wages, Firm Performance and Job Turnover.
Kallifatides, M., Modern företagsledning och omoderna företagsledare.
Kaplan, M., Acquisition of Electronic Commerce Capability - The Cases of Compaq and Dell in Sweden.
Mähring, M., IT Project Governance.
Nilsson, M., Essays in Empirical Corporate Finance and Governance.
Rekrytering av koncernstyrelsen – Nomineringsförfaranden och styrelsesammansättning med focus på kvinnors ställning och möjligheter. Sjöstrand, S-E. och Petrelius, P.,(red)
Scener ur ett företag – Organiseringsteori för kunskapssamhället. Löwstedt, J. Stymne, B.,(red).
Schenkel, A., Communities of Practice or Communities of Discipline - Managing Deviations at the Øresund Bridge.
Schuster, W., Företagets Valutarisk – En studie av horisontella och vertikala styrprocesser.
Skogsvik, S., Redovisningsmått, värderelevans och informationseffektivitet.
Sundén, D., The Dynamics of Pension Reform.
Ternström, I., The Management of Common-Pool Resources - Theoretical Essays and Empirical Evidence.
Tullberg, J., Reciprocitet – Etiska normer och praktiskt samarbete.
Westling, G., Balancing Innovation and Control – The Role of Face-to-face Meetings in Complex Product Development Projects.
Viklund, M., Risk Policy – Trust, Risk Perception, and Attitudes.
Vlachos, J., Risk Matters - Studies in Finance, Trade and Politics.

2001
Adolfson, M., Monetary Policy and Exchange Rates – Breakthrough of Pass-Through.
Andersson, P., Expertise in Credit Granting: Studies on Judgment and Decision-Making behavior.
Björklund, C., Work Motivation - Studies of its Determinants and Outcomes.

Center for Management and Organization 50 (1951-2001).

Charpentier, C., Uppföljning av kultur- och fritidsförvaltningen efter stadsdelsnämndsreformen.

Dahlén, M., Marketing on the Web - Empirical Studies of Advertising and Promotion Effectiveness.

Eckerlund, I., Essays on the Economics of Medical Practice Variations.

Ekelund, M., Competition and Innovation in the Swedish Pharmaceutical Market.

Engström, S., Success Factors in Asset Management.

Ericsson, D., Kreativitetsmysteriet – Ledtrådar till arbetslivets kreativisering och skrivandets metafysik.

Eriksson, R., Price Responses to Changes in Costs and Demand.

Frisell, L., Information and Politics.

Giordani, P., Essays in Monetary Economics and Applied Econometrics.

Gustavsson, P., Essays on Trade, Growth and Applied Econometrics.

Hedlund, A., Konsumentens erfarenhet – och dess inverkan på livsmedelsinköp på Internet.

Hill, M., Essays on Environmental Policy Analysis: Computable General Equilibrium Approaches Applied to Sweden.

Hvenmark, J., Varför slocknar elden? Om utbrändhet bland chefer i ideella organisationer.

Hägglund, P.B., Företaget som investeringsobjekt – Hur placerare och analytiker arbetar med att ta fram ett investeringsobjekt.

Höök, P., Stridspiloter i vida kjolar, om ledarutveckling och jämställdhet.

Johansson, C., Styrning för samordning.

Josephson, J., Evolution and Learning in Games.

Kjellberg, H., Organising Distribution - Hakonbolaget and the efforts to rationalise food distribution, 1940-1960.

Lange, F. och Wahlund, R., Category Management – När konsumenten är manager.

Liljenberg, A., Customer-geared competition – A socio-Austrian explanation of Tertius Gaudens.

Lindkvist, B., Kunskapsöverföring mellan produktutvecklingsprojekt.

Ljunggren, U., Nyckeltal i grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen.

Läkemedel – Kostnad eller resurs för sjukvården? Jönsson, B.,(red).

Löf, M., On Seasonality and Cointegration.

Martensen, K., Essays on Entry Externalities and Market Segmentation.

Matros, A., Stochastic Stability and Equilibrium Selection in Games.

Mårtensson, P., Management Processes – An Information Perspective on Managerial Work.

Nilsson, A., Market Transparency.

Norberg, P., Finansmarknadens amoralitet och det kalvinska kyrkorummet – En studie i ekonomisk mentalitet och etik.

Persson, B., Essays on Altruism and Health Care Markets.

Rech, G., Modelling and Forecasting Economic Time Series with Single Hidden-layer Feedforward Autoregressive Artificial Neural Networks.

Skoglund, J., Essays on Random Effects Models and GARCH.

Strand, N., Empirical Studies of Pricing.

Thorén, B., Stadsdelsnämndsreformen och det ekonomiska styrsystemet - Om budgetavvikelser.

2000

Berg-Suurwee, U., Styrning före och efter stadsdelsnämndsreform inom kultur och fritid – Resultat från intervjuer och enkät.
Bergkvist, L., Advertising Effectiveness Measurement: Intermediate Constructs and Measures.
Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Koncernchefen och ägarna.
Bornefalk, A., Essays on Social Conflict and Reform.
Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform.
Edman, J., Information Use and Decision Making in Groups.
Emling, E., Svenskt familjeföretagande.
Ericson, M., Strategi, kalkyl, känsla.
Gunnarsson, J., Wahlund, R., Flink, H., Finansiella strategier i förändring: segment och beteenden bland svenska hushåll.
Hellman, N., Investor Behaviour – An Empirical Study of How Large Swedish Institutional Investors Make Equity Investment Decisions.
Hyll, M., Essays on the Term Structure of Interest Rates.
Håkansson, P., Beyond Private Label – The Strategic View on Distributor Own Brands.
I huvudet på kunden. Söderlund, M (red). *EFI och Liber Förlag.*
Karlsson Stider, A., Familjen och firman.
Ljunggren, U., Styrning av grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen – Resultat från intervjuer och enkät.
Ludvigsen, J., The International Networking between European Logistical Operators.
Nittmar, H., Produktutveckling i samarbete – Strukturförändring vid införande av nya Informationssystem.
Robertsson, G., International Portfolio Choice and Trading Behavior.
Schwarz, B., Weinberg, S., Serviceproduktion och kostnader – att söka orsaker till kommunala skillnader.
Stenström, E., Konstiga företag.
Styrning av team och processer – Teoretiska perspektiv och fallstudier. Bengtsson, L., Lind, J., Samuelson, L.A., (red).
Sweet, S., Industrial Change Towards Environmental Sustainability – The Case of Replacing Chloroflouorocarbons.
Tamm Hallström, K., Kampen för auktoritet – standardiseringsorganisationer i arbete.

1999

Adler, N., Managing Complex Product Development.
Allgulin, M., Supervision and Monetary Incentives.
Andersson, P., Experto Credite: Three Papers on Experienced Decision Makers.
Ekman, G., Från text till batong – Om poliser, busar och svennar.
Eliasson, A-C., Smooth Transitions in Macroeconomic Relationships.
Flink, H., Gunnarsson, J., Wahlund, R., Svenska hushållens sparande och skuldsättning– ett konsumentbeteende-perspektiv.
Gunnarsson, J., Portfolio-Based Segmentation and Consumer Behavior: Empirical Evidence and Methodological Issues.
Hamrefors, S., Spontaneous Environmental Scanning.
Helgesson, C-F., Making a Natural Monopoly: The Configuration of a Techno-Economic Order in Swedish Telecommunications.
Japanese Production Management in Sunrise or Sunset. Karlsson, C., (red).
Jönsson, B., Jönsson, L., Kobelt, G., Modelling Disease Progression and the Effect of Treatment in Secondary Progressive MS. Research Report.

Lindé, J., Essays on the Effects of Fiscal and Monetary Policy.
Ljunggren, U., Indikatorer i grundskolan i Stockholms stad före stadsdelsnämnds-reformen – en kartläggning.
Ljunggren, U., En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan – Med tillämpning i Stockholms stads grundskolor.
Lundbergh, S., Modelling Economic High-Frequency Time Series.
Mägi, A., Store Loyalty? An Empirical Study of Grocery Shopping.
Mölleryd, B.G., Entrepreneurship in Technological Systems – the Development of Mobile Telephony in Sweden.
Nilsson, K., Ledtider för ledningsinformation.
Osynlig Företagsledning. Sjöstrand, S-E., Sandberg, J., Tyrstrup, M., (red).
Rognes, J., Telecommuting – Organisational Impact of Home Based – Telecommuting.
Sandström, M., Evaluating the Benefits and Effectiveness of Public Policy.
Skalin, J., Modelling Macroeconomic Time Series with Smooth Transition Autoregressions.
Spagnolo, G., Essays on Managerial Incentives and Product-Market Competition.
Strauss, T., Governance and Structural Adjustment Programs: Effects on Investment, Growth and Income Distribution.
Svedberg Nilsson, K., Effektiva företag? En studie av hur privatiserade organisationer konstrueras.
Söderström, U., Monetary Policy under Uncertainty.
Werr, A., The Language of Change The Roles of Methods in the Work of Management Consultants.
Wijkström, F., Svenskt organisationsliv – Framväxten av en ideell sektor.