TESTING THE UNIT ROOT HYPOTHESIS IN NONLINEAR TIME SERIES AND PANEL MODELS

Rickard Sandberg

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
Torsdagen den 25 November 2004
kl. 10.15 i sal Torsten
Handelshögskolan, Sveavägen 65

Testing the unit root hypothesis in nonlinear time series and panel 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 the 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. At EFI, the researchers select their 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 at the Institute are organized in 21 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 Industry Dynamics (D)	Prof Björn Axelsson
ACCOUNTING, CONTROL AND CORPORATE FINANCE	
Accounting and Managerial Finance (B)	Prof Lars Östman
Center for Financial Analysis and Managerial Economics in Accounting (BFAC)	Prof Kenth Skogsvik
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
Center for Tax Law	Prof Bertil Wiman

Chair of the Board: Professor Carin Holmquist *Director:* Associate Professor Filip Wijkström

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

TESTING THE UNIT ROOT HYPOTHESIS IN NONLINEAR TIME SERIES AND PANEL MODELS

Rickard Sandberg

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 91-7258-655-9

Keywords: Unit root, Parameter constancy, Smooth transition models, Hypothesis testing, Nonlinear panel data models

Printed by:
Elanders Gotab, Stockholm 2004

Distributed by:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm, Sweden
www.hhs.se/efi

I dedicate my thesis to my family, especially my mother, whom I miss ever so much, and my father for his enormous will power and for being the toughest fighter I know.

Contents

Acknowledgments

First of all, I would like to thank my supervisor, Changli He, for his invaluable help and knowledge. As a supervisor, he has been very patient and reassuring.

I would also like to thank my assistant supervisor, Timo Teräsvirta, for his expertise on nonlinear time series models and for being such a major source of inspiration for the content of this thesis.

I gratefully acknowledge the Jan Wallander and Tom Hedelius Foundation and Grants No. J02-35, for financial support. I am also very grateful for the contributions from the Department of Economic Statistics, SSE, which enabled me to go on unforgettable conference trips to e.g. Panama City, Texas and Madrid.

I would also like to mention the people at the Department of Economic Statistics, SSE, for being helpful and understanding.

I would of course also like to show my appreciation to my friends, Jimmy Skoglund (Jimmy "Hundred" - badminton ace from Grums), Ulf Nederman (my doubles partner from Grums), Martin Adamsson (badminton ace from Falköping), Thomas Kronberg (the goat!), Johan Nyhlén (the lawyer), Tommy Andersson (incurable hard-rocker), Sophie von Pöffendahl (missä huff?), Birgit Strikholm (Cure fan and supportive room-mate), Jan Rudander ('on-call' statistician at Karlstad University), Magnus Andersson (glamour), Per Mattson (the badminton ace), Annastiina Silvennoinen and Mika Meitz (Team Finland). Without your support, inspiration, good humour and constructive criticism, the time it has taken me to write this thesis would have been much the poorer.

My thanks finally to Jana Eklund who has helped me put this thesis together. Obliged by the newly commenced tradition, I should be tempted to put the blame on her for any mistakes or shortcomings in the thesis, but still humbly claim them solely mine.

Stockholm, October 2004
Rickard Sandberg

Part I

Summary of Thesis

Introduction and summary of the thesis

Testing the unit root hypothesis in nonlinear univariate time series models

It is well known that the traditional unit root tests based on linear models by Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988) lack power when the model under the alternative hypothesis is nonlinear (especially models with structural changes in levels and trends). This is not necessarily a drawback, however, but by empirical findings it seems that many time series exhibit nonlinear adjustment paths towards an equilibrium, see for instance Perron (1989) and Stock and Watson (1996). Evidently, in these situations the classical unit root tests are based on misspecified alternatives and are biased towards nonrejection of the null hypothesis of a unit root, as pointed out in Perron (1990). In particular this has serious implications for applied work. For example, as a macro-economist it is crucial to know how shocks (booms/recessions) to the economy will affect e.g. the GDP in the future, i.e. do the shocks have a permanent or transitory or effect, see Nelson and Plosser (1982). If the adjustment process of the GDP time series is nonlinear, then the outcomes of the traditional unit root tests are most likely fallacious and the impact of shocks to the economy is seen as permanent, when it in fact dies out over time. Obviously there is a need for unit root tests in time series models where the adjustment process towards a long-run equilibrium is nonlinear. Strong research efforts are made in this area, and most of the tests concern nonlinear models with shifts in levels and trends, see for instance Perron (1989), Perron (1990), Perron and Vogelsang (1992), Zivot and Andrews (1992), Banerjee, Lumsdaine, and Stock (1992), Leybourne, Newbold, and Vougas (1998), Harvey and Mills (2002), Saikkonen and Lütkepohl (2002) and Lanne, Lütkepohl, and Saikkonen (2003) among many others.

My contribution to this area consists of deriving several new unit root tests in nonlinear models that accommodate smooth shifts in levels, dynamics, and trends. The models considered are generalizations of the smooth transition[1] autoregressive (STAR) models by, Chan and Tong (1986), and Luukkonen, Saikkonen, and Teräsvirta (1988). As a first-order example, let $\{y\}_{t=1}^{\infty}$ be a stochastic process generated according to

$$y_t = \mathbf{x}_t'\boldsymbol{\pi}_1 + \mathbf{x}_t'\boldsymbol{\pi}_2 F(z) + u_t, \quad t = 1, ..., T, \tag{1}$$

where $\mathbf{x}_t = (1, y_{t-1})'$, $\boldsymbol{\pi}_1 = (\pi_{10}, \pi_{11})'$, $\boldsymbol{\pi}_2 = (\pi_{20}, \pi_{21})'$, $F(z)$ is the transition function discussed in detail below, and u_t is an error term. I also propose a new model by adding a time trend to the model specification in (1), in which case $\mathbf{x}_t = (1, y_{t-1}, t)'$, $\boldsymbol{\pi}_1 = (\pi_{10}, \pi_{11}, \pi_{12})'$, and $\boldsymbol{\pi}_2 = (\pi_{20}, \pi_{21}, \pi_{22})'$. This model is called the smooth transition autoregressive trend (START) model, and is a particularly interesting alternative when one wishes to test the null hypothesis of a random walk with drift. Obviously, the STAR and START models are nonlinear analogues to the linear models which Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988) base their unit root tests on.

In (1) it is the transition function $F(z)$ that captures the nonlinear adjustment process towards an equilibrium, and it is essential to be precise about the nature of $F(z)$. It is assumed that $F(z)$ satisfies the following conditions: (i) $F(0) = 0$. (ii) $F(z)$ and its derivatives up to order $n+1$ are continuous in an open interval $(-\varepsilon, \varepsilon)$ for $\varepsilon > 0$. (iii) $d^n F(z)/dz^n|_{z=0} \neq 0$ for some n. (iv) $F(z)$ is bounded for all z. Typical examples of functions belonging to this class are cumulative distribution functions (after a proper downward shift) or the trigonometric function $\sin(z)$ where $z = 2\pi t\gamma/T$ and γ denotes a particular frequency. The transition function that I adopt is the modified logistic cumulative distribution function

$$\begin{aligned} F(z) &= 1/(1+e^{-z}) - 1/2, \\ z &\equiv \gamma(t-c), \quad \gamma \geq 0, \quad c \in (0, T), \end{aligned} \tag{2}$$

and (1) with (2) is now referred to as the LSTAR and LSTART models. It is clear that (2) satisfies (i) by setting $\gamma = 0$. The conditions (ii)-(iv) are trivially met. The transition function in (2) is a non-decreasing function of time and has the range $[-1/2, 1/2]$ allowing the model to change smoothly from an initial equilibrium $E[y_t|\mathcal{F}_{t-1}] = \mathbf{x}_t'(\boldsymbol{\pi}_1 - 0.5\boldsymbol{\pi}_2)$ to a long-run equilibrium

[1] Bacon and Watts (1971) coined the term smooth transition.

$E[y_t|\mathcal{F}_{t-1}] = \mathbf{x}_t'(\boldsymbol{\pi}_1 + 0.5\boldsymbol{\pi}_2)$ with t. This emphasizes that the model (1) with (2) accommodates a smooth shift in levels, dynamics, and trends.[2] The speed of transition between the two equilibriums is determined by γ and the transition is taking place around c. Since the linear autoregressive (AR) process and the threshold autoregressive (TAR) model discussed in Chan and Tong (1986) are models frequently used as alternatives of a unit root hypothesis, it is important to note that both these models emerge as special cases of (1) with (2) by letting $\gamma \to 0$ (i.e. no transition) or $\gamma \to \infty$ (i.e. an instantaneous transition), respectively.

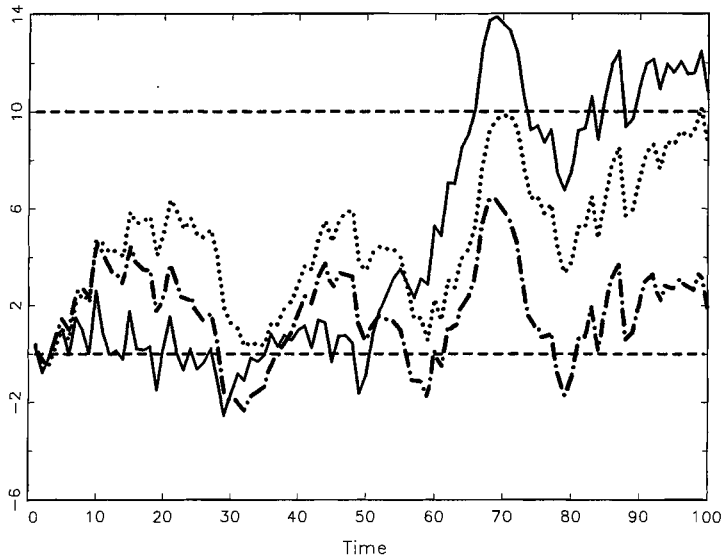
To visualize some of the nonlinear features described above, a typical realization of an LSTAR model is depicted in Figure 1. In particular a smooth shift in level is seen from the traverse from an initial equilibrium at zero towards a long-run equilibrium at 10. Figure 1 also reveals the difficulty in distinguishing between a random walk and an LSTAR model just by visual inspection, especially in small samples.

The test of a unit root in the LSTAR and LSTART models is a joint test of linearity (parameter constancy) and a unit root in the linear part. However, testing linearity, obtained by setting $\gamma = 0$ in (2), causes problems with unidentified parameters under the null hypothesis. I remedy this problem by a Taylor approximation of a suitable order of $F(t;\gamma,c)$ around $\gamma = 0$, as suggested in Luukkonen, Saikkonen, and Teräsvirta (1988). The approximation is applied to (1), and merging terms yields auxiliary regression equations that serve as a framework within which the unit root hypothesis is posed and tested. The crucial implication of the unit root hypothesis is that the asymptotic normality of the OLS estimates in the auxiliary regression equations, derived in Lin and Teräsvirta (1994), is no longer valid. Instead the distributional theory is non-standard and characterized by functionals of standard Brownian motions on the unit interval.

It should be pointed out that unit root tests already exist in the STAR framework, see for instance Enders and Granger (1998), Bec, Salem, and Carrasco (2002), Eklund (2003a), Eklund (2003b), and Kapetanios, Shin, and Snell (2003). They differ, however, fundamentally in that the transition function is assumed to be stochastic whereas in the present case it is deterministic. More closely related approaches where the transition is a function of time are found in Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002). They test a unit root in time series models with a smooth transition in the intercept and the trend, many of which are nested within the LSTAR

[2]It is not necessary to allow for a smooth transition in all parameters. One may consider a smooth transition only in a sub-set of the parameters.

Figure 1 Typical realizations of an LSTAR(1) model (solid line) with initial and long-run equilibriums (dashed line), a stationary AR(1) process (dash-dotted line), and a random walk (dotted line)



and LSTART models. More importantly, on the contrary to their simulation based results, the distributional theory is provided in detail along the testing procedures.

It is important to describe the implications of the testing philosophy that is adopted, and why it is a powerful device if one wishes to find evidence against the null hypothesis of a unit root. Since the tests are based on auxiliary regression equations, what does it mean if the null hypothesis of a non-stationary process is rejected? What can one say about the alternative? First note that essentially the same tests are obtained for any transition function with a valid Taylor approximation. Consider the following example: Applying a first-order Taylor approximation to the logistic and trigonometric functions $F(t;\gamma,c) = 1/(1+e^{-\gamma(t-c)}) - 1/2$ and $G(t;\gamma) = \sin(2\pi t\gamma/T)$, respectively, leads to the same test in the sense that both approximations yield auxiliary regressions that contain the same set of explanatory variables. Moreover, by varying the order of the approximation, tests that are nested are obtained. Consider the following example: Applying a first-order Taylor approximation

to a transition function with two transitions

$$F_2(t;\gamma,c_1,c_2) = 1/(1+e^{-\gamma(t-c_1)(t-c_2)}) - 1/2, \quad \gamma > 0, \quad c_1 < c_2,$$

in (1) yields a set of explanatory variables that are contained in the corresponding set of explanatory variables obtained by using a third-order Taylor approximation of $F(t;\gamma,c)$.

The discussion about the Taylor approximation also raises the important issue of a suitable order of approximation. This must depend upon the application, but in most of the cases, a third-order approximation of $F(t;\gamma,c)$ is enough. However, a test based on a transition function with m ($m \geq 1$) transitions

$$F_m(t;\gamma,c_1,...,c_m) = 1/(1+e^{-\gamma\Pi_{i=1}^{m}(t-c_i)}) - 1/2, \quad \gamma > 0, \quad c_1 < ... < c_m,$$

with a first-order Taylor approximation is always nested within a test based on a $(m+2)$th-order Taylor approximation of $F(t;\gamma,c)$. It seems therefore, at least conceptually, desirable to derive tests where the order of approximation of $F(t;\gamma,c)$ in (2) is arbitrary.

To conclude, the rejection of the null hypothesis is for a wide class of nonlinear models, and by increasing the order of the approximation of $F(t;\gamma,c)$, the members in the class increase. This also means that in a case of rejection the modeller is not left with a unique parametric alternative. However, suggestions and guidelines of how to proceed are for instance given in Teräsvirta (1994) and Teräsvirta (1998).

To this end, in the two first chapters in the thesis I establish the theory for a mth-order approximation of $F(t;\gamma,c)$ when testing the unit root hypothesis in pth-order LSTAR and LSTART models.

Testing the unit root hypothesis in nonlinear panel data models

In the case of a non-rejection of the unit root hypothesis in single time series by using traditional unit root tests, it has become a common practice to gather additional information by constructing a panel of single time series and to apply one of the conventional unit root tests in panels by Quah (1994), Harris and Tzavalis (1999), Levin, Lin, and Chu (2002), and Im, Pesaran, and Shin (2003). This is done in order to be able to find evidence against the null hypothesis, when for instance one examines the purchasing power parity conjecture. However, as pointed out above, many single time series seem to

be of a nonlinear nature and if one were to consider a panel, it would seem natural to consider the option that the cross sections in a panel, or at least a fraction of them, are nonlinear as well. If so, the conventional unit root tests in aforementioned papers will not have satisfactory power. Even though a panel set-up is considered, and the amount of information increased, a non-rejection of the null hypothesis is very likely. In other words, the issue of unit root tests in nonlinear panels is highly relevant.

Attempts towards this goal can be found in Im and Lee (1999), Silvestre, Barrio-Castro, and Lopez-Bazo (2001), Tzavalis (2002), and Bai and Silvestre (2003). They derive unit root tests in a panel allowing each cross section to have an instant structural shift in both the intercept and the time trend. My contribution consists of deriving tests of a common unit root hypothesis in a panel logistic smooth transition autoregressive (PLSTAR) model, where each cross section is modelled as the LSTAR model in (1) with (2). The panel thus accommodates a more gradual change in the intercepts (time-varying fixed effects, see e.g. Ahn, Lee, and Schmidt (2001) for examples) and in the dynamic structures of the cross section units over time. In Chapters 3 and 4 two different approaches to testing a common unit root in a PLSTAR model are described.

In Chapter 3 the PLSTAR model is specified under the assumption that all cross section units have the same autoregressive roots, implying that it is possible to pool the data after the fixed effects have been eliminated, and a common test of a unit root is accomplish by the least squares dummy variable (LSDV) estimator. An appealing fact is that the test statistic is derived under the assumption of a fixed time dimension and a large number of cross sections. Recall that it may be difficult in small samples to visually distinguish between a random walk and a LSTAR model. This approach is inspired by the one in Harris and Tzavalis (1999), and their results follow as special cases. The limiting distribution of the test statistic is the normal distribution and the first two moments of the test statistic are calculated analytically and inference is easily made.

One restrictive assumption about the PLSTAR model in Chapter 3 is that all cross sections under the alternative hypothesis should converge to a long-run equilibrium at the same rate. In Chapter 4 this assumption is relaxed by generalizing the PLSTAR model so that all parameters are cross section specific. In practice this means that the alternative hypothesis can be a mixture of linear, nonlinear, and non-stationary models. The test of a common unit root in the PLSTAR model is based on averaging the individual t-statistics of a unit root in a specific cross section. This approach is related to the testing

framework in linear panels by Im, Pesaran, and Shin (2003). The limiting distribution of the test statistics is the normal distribution but with moments that are hardly tractable, and as a result they are obtained by simulations.

Last, a brief summary of the main contributions in each of the four chapters follows.

Chapter 1: Testing parameter constancy in unit root autoregressive models against continuous change

In this chapter we derive tests of parameter constancy when the data generating process is non-stationary against the hypothesis that the parameters of the model change smoothly over time. To obtain the asymptotic distributions of the tests, we generalize many theoretical results, and derive new ones, in unit root econometrics. The results are derived under the assumption that the error term is a strong mixing. Small sample properties of the tests are investigated, and the results indicate that the tests have satisfactory power results when the alternative is a nonlinear model with a small shift in levels and clear change in dynamics.

Chapter 2: Dickey-Fuller type of tests against nonlinear dynamic models

In this chapter we introduce several test statistics of testing the null hypotheses of a random walk (with or without drift) against models that accommodate a smooth nonlinear shift in the level, the dynamic structure, and the trend. We derive analytical limiting distributions for all tests. Finite sample properties are examined. The performance of the tests is compared to that of the classical unit root tests by Dickey-Fuller and Phillips and Perron, and is found to be superior in terms of power.

Chapter 3: Inference for unit roots in a panel smooth transition autoregressive model where the time dimension is fixed

In this chapter we derive a unit root test against a Panel Logistic Smooth Transition Autoregressive (PLSTAR). The analysis is concentrated on the case where the time dimension is fixed and the cross section dimension tends to

infinity. Under the null hypothesis of a unit root, we show that the LSDV estimator of the autoregressive parameter in the linear component of the model is inconsistent due to the inclusion of fixed effects. The test statistic, adjusted for the inconsistency, has an asymptotic normal distribution whose first two moments are calculated analytically. To complete the analysis, finite sample properties of the test are examined. We highlight scenarios under which the traditional panel unit root tests by Harris and Tzavalis have inferior or reasonable power compared to our test.

Chapter 4: Testing unit roots in nonlinear dynamic heterogeneous panels

In this chapter we present a unit root test in a nonlinear dynamic heterogeneous panel with each cross section modelled as an LSTAR model. All parameters are viewed as cross section specific. We allow for serially correlated errors over time and heterogeneous variance among cross sections. The test is derived under three special cases: (i) The number of cross sections and observations over time are fixed. (ii) Observations over time are fixed and the number of cross sections tends to infinity. (iii) The number of observations over time tends to infinity first and the number of cross sections thereafter. Small sample properties of the test indicate modest size distortions. The power is superior to the power of the Im, Pesaran and Shin t-type of test. The results also suggest clear improvements in power compared to a univariate unit root test that allows for nonlinearities under the alternative hypothesis.

Bibliography

AHN, S. C., Y. LEE, AND P. SCHMIDT (2001): "GMM estimation of linear panel data models with time-varying individual effects," *Journal of Econometrics*, 101, 219–255.

BACON, D. W., AND D. WATTS (1971): "Estimating the transition between two intersecting straight lines," *Biometrika*, 58, 525–534.

BAI, J., AND J. SILVESTRE (2003): "Structural changes, common stochastic trends and unit roots in panel data," Unpublished manuscript.

BANERJEE, A., R. L. LUMSDAINE, AND H. J. STOCK (1992): "Recursive and sequential tests of the unit-root and trend break hypotheses: Theory and international evidence," *Journal of Business and Economic Statistics*, 10, 271–287.

BEC, F., M. B. SALEM, AND M. CARRASCO (2002): "The purchasing power parity puzzle: Evidence from a LSTAR specification," Universit de Paris I, Working paper.

CHAN, K. S., AND H. TONG (1986): "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–191.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series," *Journal of the American Statistical Association*, 74, 427–431.

EKLUND, B. (2003a): "Testing the unit root hypothesis against the logistic smooth transition autoregressive model," Stockholm School of Economics, Working paper.

——— (2003b): "A nonlinear alternative to the unit root hypothesis," Stockholm School of Economics, Working paper.

ENDERS, W., AND C. GRANGER (1998): "Unit root tests and asymmetric adjustment with an example using the term structure of interest rates," *Journal of Business and Economic Statistics*, 16, 304–311.

HARRIS, R. D. F., AND E. TZAVALIS (1999): "Inference for unit roots in dynamic panels where the time dimension is fixed," *Journal of Econometrics*, 91, 201–226.

HARVEY, D. I., AND T. C. MILLS (2002): "Unit Roots and Double Smooth Transitions," *Journal of Applied Statistics*, 29, 675–683.

IM, K. S., AND J. LEE (1999): "LM unit root test with panel data; a test robust to structural changes," Wichita State University, Working paper.

IM, K. S., M. H. PESARAN, AND Y. SHIN (2003): "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115, 53–74.

KAPETANIOS, G., Y. SHIN, AND A. SNELL (2003): "Testing for a unit root in the nonlinear STAR framework," *Journal of Econometrics*, 112, 359–379.

LANNE, M., H. LÜTKEPOHL, AND P. SAIKKONEN (2003): "Test procedures for unit roots in time series with level shifts at unknown time," *Oxford Bulletin of Economics and Statistics*, 65, 91–115.

LEVIN, A., C. LIN, AND C. CHU (2002): "Unit root tests in panel data: Asymptotic and finite-sample properties," *Journal of Econometrics*, 108, 1–24.

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

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

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

NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

PERRON, P. (1989): "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 57, 1361–1401.

——— (1990): "Testing for a unit root in a time series with a changing mean," *Journal of Business and Economic Statistics*, 8, 153–162.

PERRON, P., AND T. J. VOGELSANG (1992): "Nonstationarity and level shifts with an application to purchasing power parity," *Journal of Business and Economic Statistics*, 10, 301–320.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

QUAH, D. (1994): "Exploiting cross section variation for unit root inference in dynamic data," *Economic Letters*, 44, 9–19.

SAIKKONEN, P., AND H. LÜTKEPOHL (2002): "Testing for a unit root in a time series with level shift at unknown time," *Econometric Theory*, 18, 313–348.

SILVESTRE, C. I., T. D. BARRIO-CASTRO, AND E. LOPEZ-BAZO (2001): "Level shifts in a panel data based unit root test. An application to the rate of unemployment," Mimeo, Department of Econometrica, University of Barcelona, Spain.

STOCK, J. H., AND M. W. WATSON (1996): "Evidence on structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

TERÄSVIRTA, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218.

——— (1998): "Modeling economic relationships with smooth transition regressions," Handbook of applied economic statistics, 507-552. New York: Dekker.

TZAVALIS, E. (2002): "Structural breaks and unit roots for short panels," Queen Mary University of London, Working paper No. 459.

ZIVOT, E., AND D. W. K. ANDREWS (1992): "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.

Part II

The Chapters

Chapter 1

Testing Parameter Constancy in Unit Root Autoregressive Models Against Continuous Change

1.1 Introduction

The issue of parameter constancy in time series models is important and arises since it appears that data from many e.g. economic time series rather support models whose parameters are likely to be affected by an economical environment, see for instance Stock and Watson (1996). These time series exhibit properties that could be captured by models allowing for parameters that vary, and this most likely in a nonlinear way. Testing parameter constancy in nonlinear models becomes therefore important in attempts to detect misspecified models and avoid invalid inference.

There is vast literature on how to proceed, and the general approach is to have parameter constancy as the null hypothesis and test it against a parametric alternative. The parametric alternative and its properties are therefore in focus, and could for instance be characterized by parameters varying over time or that they change (mostly a one-time change) with some other endogenous/exogenous threshold variable, see for instance Chu and White (1991), Hansen (1992), and Andrews (1993). Time-varying parameters could be divided into two broad categories, i.e. parameters being stochastic (e.g. a random walk) or that they change according to a certain nonlinear deterministic function over time. For the former, see for instance Nyblom (1989). The latter will be in focus in this chapter.

Considering nonlinear deterministic functions, the theory has evolved from testing against abrupt changes (an instantaneous structural break) to testing against more general and flexible functions where the parameters are allowed to have multiple smooth changes over time. As such, Lin and Teräsvirta (1994) discuss testing the null hypothesis of a linear autoregressive model against the logistic smooth transition autoregression (LSTAR) model. They assume that the data generating process (DGP) is stationary under the null hypothesis and asymptotic normality holds in their framework.

However, for many time series it may be difficult just by visual inspection to distinguish between data that have been generated by a random walk or data generated by a model that is nonlinear in parameters. It is therefore natural to discuss the same issue as in Lin and Teräsvirta (1994) when the stationary assumption is relaxed, and instead consider a random walk (with or without a drift) as the DGP when the modeller tests parameter constancy in the LSTAR model. Furthermore, another appealing feature of our approach is that our test is a direct test in the sense that it is based on "raw" data, and the first step is not to make the data stationary by taking first differences.

Imposing the assumption of a random walk means that we abandon the asymptotic normality of the OLS estimates for a non-standard limiting distri-

bution characterized by various functionals of Brownian motions on the unit interval. The derived test statistic is robust for a wide class of error terms. In the sequel of finding the limiting distribution of our test, we generalize many asymptotic theoretical results in the area of unit roots that for instance could be found in Phillips (1987), Phillips and Perron (1988), and Hamilton (1994) among others, and new results are introduced.

The rest of this chapter is organized as follows. In Section 1.2 we present the parametric alternative, i.e. the LSTAR model. Joint tests of linearity and non-stationarity in an LSTAR model as well as the distributional theory of the tests and their properties are described in Section 1.3. In Section 1.4 the Monte Carlo experiments are reported. Concluding remarks are found in Section 1.5. A Lemma and proofs are given in the Mathematical Appendix.

1.2 The model

Consider the logistic smooth transition autoregression model of order p, henceforth abbreviated the LSTAR(p) model,

$$y_t = \mathbf{x}_t'\boldsymbol{\pi}_1 + \mathbf{x}_t'\boldsymbol{\pi}_2 F(t;\boldsymbol{\theta}) + u_t, \quad t = 1, ..., T, \tag{1.1}$$

where $\mathbf{x}_t' = (1,\ y_{t-1}, ..., y_{t-p})$ is a $(p+1) \times 1$ vector and $p \in \mathbb{N}$, $\boldsymbol{\pi}_1 = (\pi_{10}, ..., \pi_{1,p})' \in \mathbb{R}^{p+1}$, $\boldsymbol{\pi}_2 = (\pi_{20}, ..., \pi_{2,p})' \in \mathbb{R}^{p+1}$, u_t is an error term with properties discussed in detail in the next subsection, $F(t,\boldsymbol{\theta})$ is a logistic smooth transition function. Following Lin and Teräsvirta (1994) a full parametrization of $F(t,\boldsymbol{\theta})$ in (1.1) is given by[1]

$$F(t;\boldsymbol{\theta}) = \frac{1}{(1+\exp\{-\gamma(t^k + \alpha_1 t^{k-1} + ... + \alpha_{k-1}t + \alpha_k)\})}, \quad k = 1,2,3, \tag{1.2}$$

where $\boldsymbol{\theta} = (\gamma, \boldsymbol{\alpha}')' \in [0,\infty) \times \mathbb{R}^k$, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)'$ and $\boldsymbol{\alpha} \in \mathbb{R}^k$ such that the roots of the polynomial in (1.2) are real. For any fixed $\boldsymbol{\theta}$, $F(t)$ defines a bounded function since $F(t) : \mathbb{R}_+ \to [0,1]$. In the special case $k = 1$ and $\gamma \in (0,\infty)$ the transition is increasing in t allowing the model in (1.1) with (1.2) to change from $E[y_t|\mathcal{F}_{t-1}] \approx \mathbf{x}_t'\boldsymbol{\pi}_1$ to $E[y_t|\mathcal{F}_{t-1}] \approx \mathbf{x}_t'(\boldsymbol{\pi}_1 + \boldsymbol{\pi}_2)$ with t. The parameters γ and α_1 have a clear interpretation. The latter is a location parameter which indicates where the symmetric transition takes place, and the former determines the speed of transition. Letting $\gamma \to \infty$ implies that the transition takes place instantaneously at α_1, and F becomes the

[1] For $k = 2$, Lin and Teräsvirta (1994) define the transition function $F(t) = 1 - \exp\{-\gamma(t-\alpha)^2\}$.

indicator function: $\mathbf{1}_{(\alpha_1,0)}(t) = 1$ if $t \in (\alpha_1, 0)$ and $\mathbf{1}_{(\alpha_1,0)}(t) = 0$ if $t \notin (\alpha_1, 0)$ and the model in (1.1) becomes a threshold autoregressive (TAR) model of order p. On the other hand, letting $\gamma \to 0$ implies that $F \to 1/2$ and the resulting model in (1.1) is a linear AR(p) model with parameter vector equal to $(\boldsymbol{\pi}_1 + \boldsymbol{\pi}_2)/2$. Moreover, with $k \geq 2$ the transition function F exhibits two or more transitions and the transition paths are nonmonotonic. Finally note that for any k and $\gamma \in [0, \infty)$ we have that F is differentiable and for $\gamma = \infty$, F exhibits point(s) of discontinuity.

1.3 Testing procedures

Our aim is to test for parameter constancy in (1.1), under the assumption that the true process is a random walk. To proceed and ease up exposition, we focus on the LSTART(1) model with one transition, i.e. we let $p = k = 1$ in (1.1) and (1.2).

1.3.1 Testing parameter constancy in the LSTAR(1) model

We will first exploit the implications of linearity. We can, without loss of generality, replace F in (1.2) by a downward shift, say $\tilde{F}$, defined by

$$\tilde{F}(t, \boldsymbol{\theta}) \equiv F(t, \boldsymbol{\theta}) - 0.5. \tag{1.3}$$

It is now evident that we could test for parameter constancy in (1.1) by letting $\gamma = 0$, since $\tilde{F}(t, \boldsymbol{\theta}) = 0$ for all t and $\boldsymbol{\alpha}$. It is therefore natural to define the hypothesis $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$.[2] We have, however, an identification problem under the null hypothesis, since the vectors $\boldsymbol{\pi}_2$ and $\boldsymbol{\alpha}$ will be undetermined, see e.g. Luukkonen, Saikkonen, and Teräsvirta (1988). This problem could be solved by approximating $\tilde{F}$ in a neighborhood of $\gamma = 0$. An obvious candidate is a first-order Taylor approximation of $\tilde{F}$ around $\gamma = 0$, and yields

$$A_1(t; \gamma, \alpha_1) = 0.25\gamma(t + \alpha_1) + r_1, \tag{1.4}$$

and r_1 is the remainder. Substituting (1.4) into (1.1) gives a linear approximation of the LSTAR model,

$$y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1})\, c\gamma(t + \alpha_1) + u_t^*, \tag{1.5}$$

[2]To identify (1.1) under the alternative, one has to assume either $\gamma > 0$ or $\gamma < 0$. We choose *a priori* to rule out $\gamma < 0$ because in the defintion of F in (1.2) it is assumed that $\gamma \in [0, \infty)$.

and u_t^* is the error term adjusted with respect to the Taylor expansion where $u_t^* = u_t$ holds under the null hypothesis. Collecting the terms in (1.5) yields the reparametrized auxiliary regression model

$$y_t = \mathbf{s}_t'\boldsymbol{\lambda} + (y_{t-1}\mathbf{s}_t)'\boldsymbol{\varphi} + u_t^*, \tag{1.6}$$

where $\mathbf{s}_t = (1, t)'$, $\boldsymbol{\lambda} = (\lambda_0, \lambda_1)'$, $\boldsymbol{\varphi} = (\varphi_{01}, \varphi_{11})'$. The auxiliary null hypothesis of parameter constancy becomes,

$$H_0^{aux}: \quad \lambda_0 \in \mathbb{R}, \quad \lambda_1 = 0, \quad \varphi_{01} \in \mathbb{R}, \quad \varphi_{11} = 0. \tag{1.7}$$

The null hypothesis in (1.7) implies that the regression equation in (1.6) is reduced to an AR(1) process with an intercept since the terms involving a time trend are equal to zero.

1.3.2 Testing unit roots in the LSTAR(1) model

Lin and Teräsvirta (1994) assume that the model (1.6) is stationary under the null hypothesis, i.e. $\varphi_{01} \in (-1, 1)$. We will in contrast consider a random walk as the true DGP implying the null hypothesis $H_0 : \varphi_{01} = 1$, and assume stability under the alternative hypothesis, i.e. $H_1 : \varphi_{01} < 1$. Therefore, the joint auxiliary hypothesis of parameter constancy and non-stationarity is given by

$$H_0^{aux}: \quad \lambda_0 \in \mathbb{R}, \quad \lambda_1 = 0, \quad \varphi_{01} = 1, \quad \varphi_{11} = 0. \tag{1.8}$$

Since the model in (1.6) contains an intercept, y_{t-1} is under the null hypothesis asymptotically equivalent to $\lambda_0(t-1)$. Thereby we will have a problem with collinearity in large samples because a time trend is already included in (1.6), see e.g. Sims, Stock, and Watson (1990) and Hamilton (1994). To avoid this problem, we first note that the process under the null hypothesis is a random walk with drift implying that $E[y_t] = \lambda_0 t$. We define therefore a new explanatory variable according to $\xi_{t-1} = y_{t-1} - E_{H_0^{aux}}[y_{t-1}]$. This transformation implies the regression model

$$y_t = \mathbf{s}_t^{*\prime}\boldsymbol{\lambda}^* + \left(\xi_{t-1}\mathbf{s}_t\right)'\boldsymbol{\varphi}^* + u_t^*, \tag{1.9}$$

where $\mathbf{s}_t^* = (\mathbf{s}_t', t^2)'$, $\boldsymbol{\lambda}^* = (\lambda_0^*, ..., \lambda_2^*)'$, and $\boldsymbol{\varphi}^* = (\varphi_{01}^*, \varphi_{11}^*)' = \boldsymbol{\varphi} = (\varphi_{01}, \varphi_{11})'$. The null hypothesis of (1.8) is now equivalent to

$$H_0^{aux}: \quad \lambda_1^* \in \mathbb{R}, \quad \lambda_0^* = \lambda_2^* = 0, \quad \varphi_{01}^* = 1, \quad \varphi_{11}^* = 0. \tag{1.10}$$

All necessary transformations are now accomplished in the sense that the process under the null hypothesis of (1.10) is a random walk without drift given by $\xi_t = \xi_{t-1} + u_t$.

1.3.3 A joint test of parameter constancy and non-stationarity

We make the following general assumptions on the error term in (1.1) under the null hypothesis (1.10).

Assumption 1 *Let $\{u_t\}_{t=1}^{\infty}$ be sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying*

(A.1.1) $Eu_t = 0$ *for all* t.

(A.1.2) $\sup_{t\in\mathbb{N}} E\left|u_t\right|^{\beta+\delta} < \infty$ *for some* $\beta \in (2, \infty]$ *and* $\delta > 0$.

(A.1.3) $\sigma^2 = \lim_{T\to\infty} T^{-1} E\left(\sum_{t=1}^{T} u_t\right)^2$ *exists and* $\sigma^2 \in (0, \infty)$.

(A.1.4) $\{u_t\}_{t=1}^{\infty}$ is strong mixing with mixing coefficients $\alpha(m)$ satisfying $\sum_{m=1}^{\infty} \alpha(m)^{1-2/\beta} < \infty$.

Under Assumption 1 the limiting distribution for the OLS estimators $(\hat{\boldsymbol{\lambda}}^{*\prime}, \hat{\boldsymbol{\varphi}}^{*\prime})$ in (1.9) under the null of (1.10) can be derived. The result is given in the following theorem.

Theorem 1.1 *Consider model (1.9) when (1.10) holds. Furthermore, assume that $\{u_t\}_{t=1}^{\infty}$ satisfies Assumption 1. Then the least square estimator $\hat{\boldsymbol{\psi}} = \left(\sum_{t=1}^{T} \mathbf{h}_t\mathbf{h}_t'\right)^{-1}\left(\sum_{t=1}^{T} \mathbf{h}_t y_t\right)$ of $\boldsymbol{\psi} = (\boldsymbol{\lambda}^{*\prime}, \boldsymbol{\varphi}^{*\prime})'$ in (1.9), where $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\lambda}}^{*\prime}, \hat{\boldsymbol{\varphi}}^{*\prime})'$ and $\mathbf{h}_t = (\mathbf{s}_t^{*\prime}, (\xi_{t-1}\mathbf{s}_t)')'$, has the following asymptotic properties*

$$\gamma_T(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} \boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}, \tag{1.11}$$

$$\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \xrightarrow{p} \mathbf{0}, \tag{1.12}$$

where $\gamma_T = diag\left\{\ T^{1/2} \quad T^{3/2} \quad T^{5/2} \quad T \quad T^2\ \right\}$, and

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{M}_1 & \sigma\mathbf{M}_2 \\ \sigma\mathbf{M}_2' & \sigma^2\mathbf{M}_3 \end{bmatrix}, \quad \boldsymbol{\Pi} = \begin{bmatrix} \sigma\boldsymbol{\Pi}_1 \\ (\sigma^2/2)\,\boldsymbol{\Pi}_2' \end{bmatrix},$$

where the sub-matrices are given by

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix},$$

$$\mathbf{M}_2 = \begin{bmatrix} \int_0^1 W(r)dr & \int_0^1 rW(r)dr \\ \int_0^1 rW(r)dr & \int_0^1 r^2W(r)dr \\ \int_0^1 r^2W(r)dr & \int_0^1 r^3W(r)dr \end{bmatrix},$$

$$\mathbf{M}_3 = \begin{bmatrix} \int_0^1 W^2(r)dr & \int_0^1 rW^2(r)dr \\ \int_0^1 rW^2(r)dr & \int_0^1 r^2W^2(r)dr \end{bmatrix},$$

$$\begin{aligned}
\mathbf{\Pi}_1 &= \begin{bmatrix} W(1) \\ W(1) - \int_0^1 W(r)dr \\ W(1) - 2\int_0^1 rW(r)dr \end{bmatrix}, \\
\mathbf{\Pi}_2 &= \begin{bmatrix} W(1)^2 - \bar{\sigma}_u^2/\sigma^2 \\ W(1)^2 - \int_0^1 W^2(r)dr - \bar{\sigma}_u^2/(2\sigma^2) \end{bmatrix},
\end{aligned}$$

where $\xrightarrow{p}$, *and* $\xrightarrow{d}$ *denotes convergence in probability and distribution, respectively,* $W(r)$ *abbreviates a standard Brownian motion on* $[0,1]$, *and* $\bar{\sigma}_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T Eu_t^2$ *such that* $\bar{\sigma}_u^2 \in (0,\infty)$.

Proof. See Appendix A. ■

Note that under the null hypothesis (1.10), $\varphi_{01}^* = \pi_{11}$. It follows from Theorem 1.1 that the limiting distribution of $T(\hat{\varphi}_{01}^* - 1)$ is given as follows.

Corollary 1.2 *Suppose that the conditions of Theorem 1.1 hold. Define the test statistic*

$$T_n \equiv T(\hat{\varphi}_{01}^* - 1).$$

Then

$$T_n \xrightarrow{d} Q_1\left(W(r)\right) + Q_2\left(W(r); \sigma^2, \bar{\sigma}_u^2\right), \tag{1.13}$$

where Q_1 *and* Q_2 *are functions of Brownian motions defined in the Appendix A. Remark that* Q_2 *depends upon the nuisance parameters* σ^2 *and* $\bar{\sigma}_u^2$. *Hence, define the adjusted restricted test statistic*

$$T_a \equiv T_n - \hat{Q}_2,$$

where

$$\hat{Q}_2 = T^2\left(\sigma^2 - \bar{\sigma}_u^2\right)\Lambda_{44}/\left(2\sigma^2 s_T^2\right) + T^3\left(\sigma^2 - \bar{\sigma}_u^2\right)\Lambda_{45}/\left(4\sigma^2 s_T^2\right),$$

with

$$\begin{aligned}
\Lambda_{44} &= s_T^2 \mathbf{r}_1 \left[\sum\nolimits_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1} \mathbf{r}_1', \quad \mathbf{r}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \\
\Lambda_{45} &= s_T^2 \mathbf{r}_1 \left[\sum\nolimits_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1} \mathbf{r}_2', \quad \mathbf{r}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\
s_T^2 &= \sum_{t=1}^T (y_t - \mathbf{h}_t'\hat{\psi}_1)^2/(T-5).
\end{aligned}$$

Then,

$$T_a \xrightarrow{d} Q_1\left(W(r)\right). \tag{1.14}$$

Proof. See Appendix A. ■

There are several things to note about the results in (1.13) and (1.14). First, both tests are invariant with respect to λ_0 in (1.8), meaning that both tests are invariant with respect to if the true model is a random walk with or without drift. Second, it is only the limiting distribution for T_a that is invariant with respect to all nuisance parameters $\{\lambda_0, \sigma^2, \bar{\sigma}_u^2\}$, whereas, as shown in appendix, the limiting distribution for T_n is a function of σ^2 and $\bar{\sigma}_u^2$. However, if $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence, $\sigma^2 = \bar{\sigma}_u^2$ holds, and it follows that $T_n \xrightarrow{d} Q_1$. Third, even if the limiting distribution Q_1 is nuisance parameter free, the test statistic T_a itself contains nuisance parameters. This means that we must replace σ^2 and $\bar{\sigma}_u^2$ with some consistent estimates to operationalize the test, see for instance Phillips (1987). So far only the properties under the null hypothesis are mentioned, but under the alternative hypothesis and especially that $\varphi_{01}^* \in (-1, 1)$, the estimates $\hat{\psi}$ from the auxiliary regression equation in (1.9) are normally distributed, see Lin and Teräsvirta (1994). In particular, in our case it follows that $\hat{\varphi}_{01}^*$ is $\mathcal{O}_p(T)$ rather than $\mathcal{O}_p(T^{1/2})$.

Furthermore, it is natural to consider the F_{OLS}-test of the joint hypothesis in (1.10) which can be expressed as $\mathbf{R}\psi = \mathbf{r}$, where $\mathbf{R} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_4 \end{bmatrix}$ and $\mathbf{0}$ is the 4×1 the null vector and $\mathbf{r} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}'$. The F_{OLS}-test statistic and its limiting distribution are given in the following corollary.

Corollary 1.3 *Define the OLS F statistic*

$$F_{OLS} \equiv \left(\mathbf{R}\hat{\psi} - \mathbf{r}\right)' \left\{ S_T^2 \mathbf{R} \left[\sum_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t' \right]^{-1} \mathbf{R}' \right\}^{-1} \left(\mathbf{R}\hat{\psi} - \mathbf{r}\right) / 4. \tag{1.15}$$

Consider model (1.9) when (1.10) holds. Assume that $\{u_t\}_{t=1}^{\infty}$ satisfies Assumption 1. Then

$$F_{OLS} \xrightarrow{d} \mathbf{\Pi}'\mathbf{\Psi}^{-1}\mathbf{R}' \left\{ \bar{\sigma}_u^2 \mathbf{R}\mathbf{\Psi}^{-1}\mathbf{R}' \right\}^{-1} \mathbf{R}\mathbf{\Psi}^{-1}\mathbf{\Pi}/4. \tag{1.16}$$

Furthermore, if $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence, then,

$$F_{OLS} \xrightarrow{d} \begin{bmatrix} \mathbf{\Pi}_1 \\ \mathbf{\Pi}_2/2 \end{bmatrix}' \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \mathbf{R}' \left\{ \mathbf{R} \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \mathbf{R}' \right\}^{-1}$$
$$\times \mathbf{R} \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Pi}_1 \\ \mathbf{\Pi}_2/2 \end{bmatrix} / 4. \tag{1.17}$$

Proof. See Appendix A. ■

Without the i.i.d. assumption for the error term, the limiting distribution in (1.16) of F_{OLS} would contain the nuisance parameters σ^2 and $\bar{\sigma}_u^2$. However, adding the i.i.d. assumption gives that $\sigma^2 = \bar{\sigma}_u^2$ holds and it is immediately seen that (1.17) is nuisance parameter free.

1.4 Monte Carlo experiments

1.4.1 Asymptotic critical values and empirical size of the parameter constancy tests

In this subsection we present the critical values for the adjusted restricted test statistic, T_a, in Corollary 1.2. For illustration and comparison we also present the critical values for the corresponding test statistic based on a third-order Taylor expansion denoted T_3.[3] The critical values for the F_{OLS} test statistic in Corollary 1.3 are presented as well. When generating the asymptotic distributions, we let $T = 1\ 000\ 000$ to simulate a Brownian motion $W(r)$ on $[0, 1]$, and the number of replications are set to 1 000 000, yielding the asymptotic critical values. The finite-sample critical values are obtained by simulating data from the model $y_t = y_{t-1} + u_t$ where $u_t \sim nid(0, 1)$ with desired sample sizes, and thereafter the test statistics in (1.14) and (1.15) are calculated. This procedure is repeated 1 000 000 times, yielding the finite-sample distributions of the tests. The results are shown in Table 1.1.

In Table 1.1 we see that the rejection of the null hypothesis requires large absolute values when using the T_a or T_3 tests, and the empirical distributions are heavily skewed to the left. We also see that the critical values for the T_a test converge faster to the asymptotic critical values than for the T_3 test, but both tests and their critical values at small sample sizes provide rather poor approximations of the asymptotic critical values. For the F_{OLS} test we see that it requires substantially higher observed values to reject the null hypothesis than compared to the standard F_{OLS} distribution.

In the case $u_t \sim nid(0, 1)$ we can ignore the estimation of the nuisance parameters σ^2 and $\bar{\sigma}_u^2$ and we have that $\hat{Q}_2 = 0$ holds for all sample sizes. However, allowing for some more complex structure for the error term, we have to estimate $\hat{Q}_2$. It is well known that classical unit root tests suffer from rather large size distortions when the error term follows an MA(1) process with moving average coefficient close to plus/minus 1, see e.g. Schwert (1989)

[3]The critical values for T_3 in Table 1.1 refer to the test statistic $T_3 = T(\hat{\varphi}_{01}^* - 1)$, based on the auxiliary regression $y_t = \mathbf{s}_t^{*\prime}\boldsymbol{\lambda}^* + (\xi_t \mathbf{s}_t)'\boldsymbol{\varphi}^* + u_t^*$, where $\mathbf{s}_t^* = (s_t', t^4)'$, $\boldsymbol{\lambda}^* = (\lambda_1^*, ..., \lambda_5^*)$, $\boldsymbol{\varphi}^* = (\varphi_{01}^*, ..., \varphi_{04}^*)'$, and u_t^* is an error term with the same properties as the error term in (1.1).

Table 1.1 Critical values for the parameter constancy tests when the DGP is a random walk with i.i.d. increments.

	Probability that T_a is less than entry			Probability that T_3 is less than entry			Probability that F_{OLS} is less than entry		
T	0.01	0.05	0.10	0.01	0.05	0.10	0.90	0.95	0.99
50	-44.28	-32.82	-27.05	-119.64	-90.06	-75.56	4.61	5.38	7.12
100	-49.33	-35.70	-29.18	-138.48	-103.72	-86.31	4.42	5.10	6.51
250	-52.98	-37.69	-30.59	-159.44	-116.44	-95.79	4.30	4.91	6.24
500	-54.83	-38.72	-31.34	-168.48	-121.80	-99.73	4.28	4.88	6.16
1000	-54.83	-38.72	-31.34	-173.38	-124.70	-101.80	4.26	4.87	6.13
∞	-55.62	-39.35	-31.87	-175.31	-126.50	-103.11	4.26	4.85	6.09

Notes: The probability shown at the head of each column is the area in the left-hand tail. The results are based on 1 000 000 replications.

and Hall (1989). We choose therefore to study the empirical size of the T_a test in (1.14) when the error term is given by the MA(1) process $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$ with $\varepsilon_t \sim nid(0,1)$ and $\theta \in \{-0.8, -0.5, 0, 0.5, 0.8\}$. To see how this affects the size of the T_a test, we calculate the empirical size (estimated size) for $T = 50, 100, 250$ and 1 000, based on 1 000 000 replications. The empirical size, for a fixed sample size and significance level, is measured by calculating $T_n - \hat{Q}_2$ and seeing how many times it is less than the critical value given in Table 1.1. To operationalize $\hat{Q}_2$ we use the Newey-West estimator with lag truncation parameter l equal to 4 to estimate σ^2, and $\bar{\sigma}_u^2$ is estimated, consistently, with $T^{-1}\sum \hat{u}_t^2$.[4] The results are presented in Table 1.2.

In Table 1.2 the empirical sizes reported in the first columns are based on the unadjusted test (UAT) statistic , i.e. we don't subtract $\hat{Q}_2$ from Q_1. The results in the second columns are based on $T_n - \hat{Q}_2$, i.e. the adjusted test (AT) statistic. From this table we see that we reject the null hypothesis far too often for negative values of θ, and also that the test is undersized for positive values of θ. Of course one hopes that AT statistic would perform significantly better than the UAT statistic, and with an estimated size that is close to the nominal size. This is not the case, however, and the improvement upon the empirical size is rather modest. It is clear that the adjustment is far from

[4]The lag-truncation parameter in the Newey-West estimator is in our case chosen somewhat arbitrarily. Phillips (1987) shows that the Newey-West estimator can provide a consistent estimate of σ^2 in an MA(∞) process when l, $T \to \infty$ such that $l/T^{1/4} \to 0$ holds.

Table 1.2 The empirical size of T_1 when the DGP is a random walk with MA(1) increments.

	T=50		T=100		T=250		T=1000	
	UAT	AT	UAT	AT	UAT	AT	UAT	AT
θ								
-0.8	0.845	0.695	0.982	0.976	0.994	0.992	0.981	0.946
-0.5	0.463	0.444	0.597	0.551	0.672	0.439	0.695	0.050
0.0	0.050	0.053	0.050	0.051	0.050	0.051	0.050	0.050
0.5	0.003	0.013	0.003	0.014	0.003	0.015	0.003	0.014
0.8	0.002	0.006	0.001	0.007	0.001	0.007	0.001	0.007

Notes: The nominal size is 5% and the results are based on 10 000 replications. UAT refers to (1.14) without the adjusting term. AT abbreviates the test in (1.14) with the adjusting term.

satisfactory. These results are in line with the findings for the classical unit root tests examined in Schwert (1989).

The poor size properties for the adjusted test statistic are explained by the fact that the adjustment factor $\hat{Q}_2$ is hardly correct in small samples. However, even for $T = 1000$ the estimates of σ^2 and $\bar{\sigma}^2$ can be quite far away from their true values. This problem could be solved, at least partially, by considering the approach of instrumental variables as in Hall (1989) or by adding lags of the error term. In our case, the latter would be equivalent to considering an LSTAR(p) model instead of the LSTAR(1) model. Schwert (1989) reports that for the classical augmented Dickey-Fuller unit root tests with 12 lags included and the corresponding t-type of test statistic, the size distortions are eliminated. Even though the deficit performance of the adjustment factor $\hat{Q}_2$, Campbell and Perron (1991) argue that a false rejection of the kind reported above is not necessarily bad, and in fact when $\theta \approx -1$, we essentially reject for a white noise process. To this end, we also studied the size properties of the T_a test, however not reported here, under the assumption that the error term was an AR(1) process. Also in this case it is found that the finite-sample performance of the adjustment quantity $\hat{Q}_2$ is non-satisfactory, and the test becomes undersized. In general it seems that the Newey-West estimator of σ^2 is hardly precise enough, and even though we in theory allow for possibly weakly dependent heterogeneously distributed errors, we may in practice end up in situations where the estimated size is either very high or low.

1.4.2 Power of the parameter constancy tests

We continue with investigation of the power properties of the parameter constancy tests. The DGP is chosen to be an LSTAR(1) model with one transition, i.e. $k = 1$. The experiments are conducted for the sample sizes $T = 100$ and $T = 250$, and we let $u_t \sim nid(0, 1)$. The stability restrictions $\pi_{11} \in (0, 1)$ and $\pi_{11} + \pi_{21} \in (0, 1)$ are imposed on the skeleton of the LSTAR(1) to rule out non-stable or explosive trajectories under the alternative hypothesis. Specifically, the autoregressive coefficients are assigned values according to the two scenarios

$$\text{Scenario I:} \quad \pi_{11} \in \{0.65, 0.70, 0.75, 0.80, 0.85\}, \quad \pi_{21} = 0.1;$$

$$\text{Scenario II:} \quad \pi_{11} = 0.1, \quad \pi_{21} \in \{0.65, 0.70, 0.75, 0.80, 0.85\}.$$

The remaining parameters are chosen to be the same in the two scenarios and are given by

$$\pi_{10} = 0, \quad \pi_{20} = 1, \quad \gamma \in \{0.01,\ 1,\ 100\}, \quad \alpha_1 = -T/2.$$

In Scenario I the autoregressive parameter varies in the linear part while it is kept fixed in the nonlinear part. Vice versa is exemplified in Scenario II. Specifically, the time series realized by Scenario I displays autoregressive parameters that at the beginning of the period are in a range of relatively high values. As time evolves, a nonlinear adjustment towards a new level with an autoregressive parameter almost equal to the autoregressive parameter at the beginning of the period takes place That is, the difference between the two autoregressive parameters at the beginning and at the end of the period is about 0.1, depending on if a complete transition takes place or not. By Scenario II we realize time series with a low value of the autoregressive parameter at the beginning of the period, and here a nonlinear adjustment towards a new level implies that the autoregressive parameter lie in a range of high values at the end of the period. The differences in the autoregressive parameters between the beginning and the end of the period range now from about 0.55 to 0.75 (depending on if a complete transition takes place or not). Scenario I represents less nonlinear time series because the impact of a nonlinear change in dynamics is suppressed. Consequently, Scenario II reveals more nonlinear time series since the nonlinear changes in the dynamics are more pronounced. The difference between the intercept in the linear and nonlinear part in the LSTAR(1) model is set modest and equals 1. Furthermore, the influence of the speed of transition, occurring at time $T/2$, between regimes

by varying γ is studied. With $\gamma = 0.01$ the LSTAR model is almost linear, and this is the only case when a complete transition does not take place over the sample periods because $F(t = T = 100; \gamma = 0.01, \alpha_1 = -50) \approx 0.62$ and $F(t = T = 250; \gamma = 0.01, \alpha_1 = -125) \approx 0.78$. Letting $\gamma = 1$ implies a smooth transition from one regime to another, and finally letting $\gamma = 100$ means that an instant single structural break between two extreme regimes takes place, and the result is a threshold autoregressive model.

One aim with Scenarios I and II is to illustrate the well-known fact that the tests based on first-order Taylor approximation suffer from low power when only a change in the intercept is considered, see Luukkonen, Saikkonen, and Teräsvirta (1988). We expect therefore that Scenario I will generate quite poor power results for the T_a test since the main source of power for this test would come from a nonlinear change in the dynamics – a feature that is suppressed in Scenario I. The opposite is expected in Scenario II – here a more evident nonlinear change in the dynamics takes place and we would expect more convincing power results for T_a. To this end we note that by the design of the experiments above we realize time series that are initially stable with autoregressive parameter π_{11} and having an attractor at $\pi_{10} = 0$. A nonlinear adjustment towards a new long-run equilibrium, i.e. the attractor $\pi_{20}(/1 - \pi_{11} - \pi_{21}) \in [4, 20]$,[5] characterized by the autoregressive parameter $\pi_{11} + \pi_{21}$, takes place. Time series generated under Scenarios I and II have the same attractors, but in the way the trajectories travel from the one attractor to the other is different. Note also that the new level of the long-run equilibrium does not only depend on π_{20} but also π_{11} and π_{12}, a property that we choose to call the level leverage effect. We expect the leverage effect to have negative influence on the power for a test based on a first order approximation, because when $\pi_{11} + \pi_{12}$ is high, the difference in levels between the beginning and the end of the sample becomes large even for small values of π_{20}. In addition, when $\pi_{11} + \pi_{12}$ is close to unity, we encounter a process that is close to a unit root process at the end of the period, which also causes a reduction in power.

The power analysis is restricted to the tests T_a and T_3 in order to illuminate the difference between a first and third-order Taylor approximation. Moreover, to our knowledge, there are no similar parameter constancy tests (with a non-stationary null hypothesis), but we note that the auxiliary regression equation in (1.9) nests the models $y_t = a_0 + a_1 y_{t-1} + u_t$ and $y_t = a_0 + a_1 y_{t-1} + a_2 t + u_t$, considered in Dickey and Fuller (1979) and Phillips and Perron (1988), and that their null hypotheses coincide with ours. The unit root test based on the former model will be denoted the DF test, and the test based on the latter

[5] Assuming that a complete transition takes place.

Table 1.3 Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario I.

$T = 100$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{11}								
	0.65	0.534	0.697	0.999	0.999	0.981	0.997	0.108	0.203
	0.70	0.414	0.583	0.987	0.998	0.912	0.975	0.087	0.159
$\gamma = 0.01$	0.75	0.297	0.454	0.903	0.974	0.734	0.884	0.067	0.137
	0.80	0.192	0.324	0.632	0.822	0.465	0.664	0.057	0.120
	0.85	0.114	0.207	0.239	0.420	0.222	0.380	0.052	0.099
	0.65	0.337	0.494	0.190	0.454	0.722	0.877	0.229	0.355
	0.70	0.196	0.327	0.024	0.097	0.425	0.637	0.207	0.331
$\gamma = 1.00$	0.75	0.084	0.164	0.001	0.005	0.158	0.303	0.200	0.320
	0.80	0.020	0.048	0.000	0.000	0.028	0.071	0.182	0.310
	0.85	0.003	0.008	0.000	0.000	0.001	0.002	0.163	0.275
	0.65	0.323	0.483	0.185	0.447	0.702	0.871	0.301	0.368
	0.70	0.187	0.317	0.023	0.090	0.409	0.622	0.223	0.340
$\gamma = 100$	0.75	0.079	0.157	0.001	0.005	0.146	0.293	0.205	0.322
	0.80	0.019	0.046	0.000	0.000	0.025	0.065	0.187	0.315
	0.85	0.002	0.007	0.000	0.000	0.000	0.002	0.172	0.293

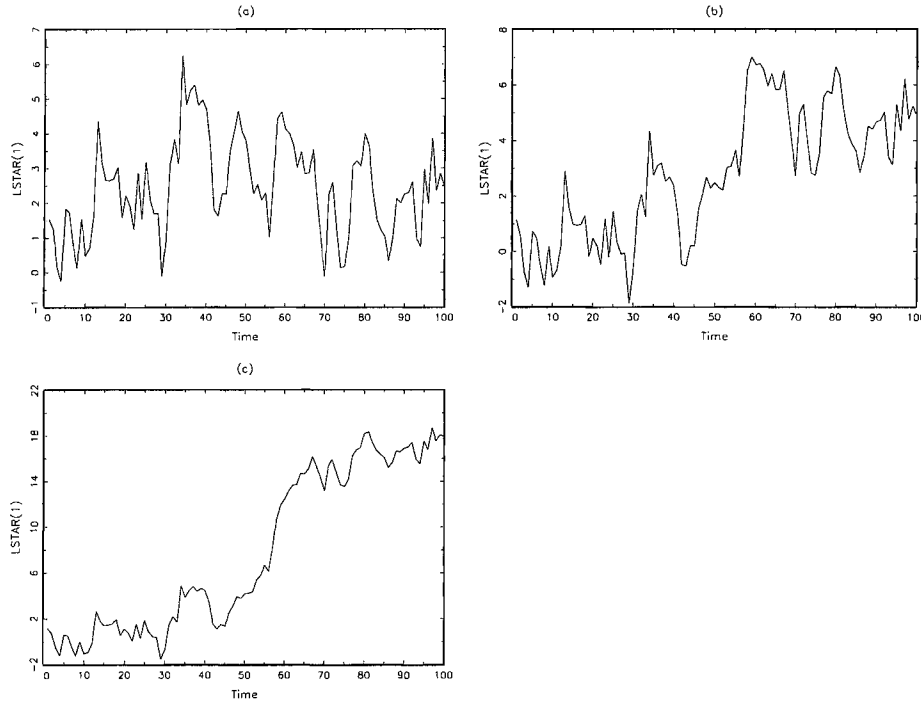
Notes: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

model will henceforth be abbreviated to the PP test. The power results for the tests under Scenario I are found in Tables 1.3 and 1.4, and for Scenario II in Tables 1.5 and 1.6.

It is seen in Table 1.3 that the DF test performs uniformly the best among all tests when $\gamma = 0.01$. This is to be expected since the LSTAR model appears linear without any evidence of a time trend or other pronounced nonlinearities, see panel (a) in Figure 1.1 (an LSTAR(1) model with $\gamma = 0.01$, $\pi_{11} = 0.7$, and $\pi_{21} = 0.1$). The power of the T_a and T_3 test statistics are modest because when the LSTAR model is nearly linear, the many regressors come to no efficient use and the tests are instead penalized. By this reasoning it is clear why the power for T_3 is lower than that for T_a which in turn is lower than that for PP.

Increasing the speed of transition to $\gamma = 1.00$, we see a turn in ranking among the test. The DF test is now inferior and its power is close to zero or

Figure 1.1 Realizations of the LSTAR(1) models under Scenario I.



equal to zero for high values of π_{11}. We see that the PP test still performs better than the T_a test. In fact, the PP test has substantial power for low values of π_{11} which may be explained by that the trajectories display rather linear time trends, see panel (b) in Figure 1.1 (an LSTAR(1) model with $\gamma = 1$, $\pi_{11} = 0.7$, and $\pi_{21} = 0.1$). The presence of a possible time trend and/or shift in levels also explains the remarkable drop in power for the DF test. The non-satisfactory performance of the T_a test just confirms the above stated expectations under Scenario I. If a complete transition takes place the impact of the level leverage effect on the power for T_a, by increasing π_{11}, without a clear nonlinear change in dynamics is revealed. The T_3 test is robust in the sense that it is quite independent of the values of π_{11}, and is in fact the only test that has power for high values of π_{11}. This can also be understood by inspection of panel (c) in Figure 1.1 (an LSTAR(1) model with $\gamma = 1.00$, $\pi_{11} = 0.85$, and $\pi_{21} = 0.1$), where the change in level between the regimes at the beginning and the end of the sample period is about 20 units. To this end

Table 1.4 Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario I.

$T = 250$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{11}								
	0.65	0.992	0.998	1.000	1.000	1.000	1.000	0.410	0.568
	0.70	0.971	0.990	1.000	1.000	1.000	1.000	0.324	0.473
$\gamma = 0.01$	0.75	0.912	0.961	1.000	1.000	1.000	1.000	0.248	0.378
	0.80	0.766	0.877	0.976	0.997	0.998	1.000	0.171	0.295
	0.85	0.512	0.680	0.291	0.531	0.862	0.949	0.119	0.210
	0.65	0.933	0.971	1.000	1.000	1.000	1.000	0.652	0.788
	0.70	0.793	0.894	0.875	0.991	0.999	1.000	0.596	0.731
$\gamma = 1.00$	0.75	0.477	0.656	0.072	0.356	0.899	0.977	0.526	0.675
	0.80	0.100	0.208	0.000	0.000	0.222	0.444	0.488	0.648
	0.85	0.007	0.001	0.000	0.000	0.000	0.001	0.433	0.600
	0.65	0.930	0.970	1.000	1.000	1.000	1.000	0.654	0.793
	0.70	0.787	0.890	0.874	0.991	0.999	1.000	0.593	0.737
$\gamma = 100$	0.75	0.472	0.648	0.070	0.356	0.896	0.976	0.529	0.695
	0.80	0.099	0.203	0.000	0.000	0.217	0.440	0.485	0.649
	0.85	0.006	0.001	0.000	0.000	0.000	0.002	0.429	0.610

Notes: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

we note that T_3 is the only test whose power is increasing in γ. The same conclusions as with $\gamma = 1.00$ hold for $\gamma = 100$, and the power results for all tests are hardly affected by a more instantaneous transition.

We continue examining Scenario I with a larger sample size $T = 250$. From Table 1.4 we see that for $\pi_{11} \leq 0.75$ the power of the T_a increases to quite satisfactory levels, and that the PP test has the highest power among all tests. Even though the sample size is increased, we see that the impact of the level leverage effect on the T_a test still dominates, and as soon as a complete transition takes places the power for T_a is rapidly decreasing in π_{11}. The power for the T_3 test for $\gamma \geq 1.00$ is substantial and quite robust against variations in π_{11}. Once again it is the only test having reasonable power with high values of π_{11}.

Table 1.5 Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario II.

$T = 100$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{21}								
	0.65	0.970	0.989	1.000	1.000	1.000	1.000	0.307	0.450
	0.70	0.965	0.988	1.000	1.000	1.000	1.000	0.285	0.442
$\gamma = 0.01$	0.75	0.956	0.983	1.000	1.000	1.000	1.000	0.267	0.416
	0.80	0.950	0.977	1.000	1.000	1.000	1.000	0.254	0.411
	0.85	0.938	0.976	1.000	1.000	1.000	1.000	0.240	0.390
	0.65	0.976	0.989	0.538	0.825	0.959	0.988	0.661	0.789
	0.70	0.938	0.975	0.134	0.453	0.854	0.939	0.715	0.839
$\gamma = 1.00$	0.75	0.833	0.912	0.007	0.034	0.559	0.755	0.785	0.883
	0.80	0.489	0.636	0.000	0.000	0.133	0.290	0.865	0.936
	0.85	0.071	0.125	0.000	0.000	0.001	0.005	0.917	0.964
	0.65	0.970	0.983	0.519	0.817	0.959	0.988	0.667	0.807
	0.70	0.932	0.967	0.123	0.341	0.828	0.932	0.729	0.851
$\gamma = 100$	0.75	0.814	0.893	0.006	0.027	0.536	0.722	0.806	0.900
	0.80	0.462	0.614	0.000	0.000	0.123	0.278	0.884	0.943
	0.85	0.056	0.106	0.000	0.000	0.001	0.005	0.932	0.971

Notes: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

From Table 1.5, illustrating Scenario II with $T = 100$, it is observed that for $\gamma = 0.01$ the classical unit root tests still have better power due to the linear properties of the LSTAR model. What is interesting is the positive change in power for the T_a test compared to Table 1.3. It is obvious that the presence of more pronounced nonlinear changes in dynamics result in an increase in power. For an example, the power in the case ($\pi_{11} = 0.7, \pi_{21} = 0.1$) of Table 1.3 equals 0.196, whereas the power in and the case ($\pi_{11} = 0.1, \pi_{21} = 0.7$) of Table 1.5 equals 0.938. The increase also confirms the expectations for Scenario II. In fact, it is the explanatory variable $t\xi_{t-1}$ in the auxiliary regression (1.9) that accounts for the increase in power compared to Scenario I. The power for the T_3 test has also increased, but suffers still from loss in power due to its ambiguity in modelling a relatively simple DGP. The power results of unity for the DF and PP tests are also rather expected since the autoregressive coefficient never

exceeds 0.63 during the whole sample (which should be compared to that the autoregressive coefficient Scenario I never exceeds 0.92, explaining why the power for the DF and PP tests is lower in Table 1.3 than in Table 1.5).

Table 1.6 Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario II.

$T = 250$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{21}								
	0.65	1.000	1.000	1.000	1.000	1.000	1.000	0.918	0.964
	0.70	1.000	1.000	1.000	1.000	1.000	1.000	0.913	0.957
$\gamma = 0.01$	0.75	1.000	1.000	1.000	1.000	1.000	1.000	0.911	0.952
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	0.898	0.950
	0.85	1.000	1.000	1.000	1.000	1.000	1.000	0.897	0.950
	0.65	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.999
	0.70	1.000	1.000	0.998	1.000	1.000	1.000	0.999	1.000
$\gamma = 1.00$	0.75	1.000	1.000	0.527	0.896	0.998	1.000	1.000	1.000
	0.80	0.983	0.996	0.000	0.001	0.754	0.906	1.000	1.000
	0.85	0.142	0.285	0.000	0.000	0.002	0.016	1.000	1.000
	0.65	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.999
	0.70	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
$\gamma = 100$	0.75	1.000	1.000	0.529	0.897	0.998	1.000	1.000	1.000
	0.80	0.980	0.994	0.000	0.009	0.740	0.890	1.000	1.000
	0.85	0.130	0.265	0.000	0.000	0.002	0.018	1.000	1.000

Notes: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

The benefits from including nonlinear regressors become clearer when studying the cases $\gamma \geq 1.00$. We see that the T_a test now performs uniformly better than the PP and DF tests at all values of π_{21}. This is plausible since none of the DF and PP tests contains the regressor $t\xi_{t-1}$ combined with the fact that a full transition takes place. Furthermore, the power for T_a is decreasing with γ and π_{21} which indicates the trade-off between the fact that a distinct difference in π_{11} and π_{21} is needed to gain power through the term $t\xi_{t-1}$, and the fact that the larger the difference in π_{11} and π_{21} is, the more pronounced the level leverage effect will be. The domination of the latter is, however, less evident than in Scenario I. Moreover, the opposite holds for the T_3 test where

the power actually increases with γ and π_{21}. That is, with evident changes in levels and dynamics, the advantages with a third-order Taylor expansion are illuminated.

The results when increasing the sample size under Scenario II are presented in Table 1.6. We see that the same conclusions can be drawn as from Table 1.5. However, we note that the power for T_a is close to unity, except for models that are almost non-stable at the end of the period, which holds independently of the value of γ. Moreover, the T_3 test exhibits a rather extraordinary robustness in power close to unity regardless of how γ and π_{21} are varied. Especially in a larger sample, the many regressors come into their own and a test based on third-order Taylor approximation seems highly preferable.

1.5 Conclusions

In this chapter we derive tests for parameter constancy in an first order LSTAR model when the null hypothesis is a random walk (possibly with a drift). This means that we relax the stationarity assumption made by Lin and Teräsvirta (1994) under the null hypothesis in the LSTAR model. We argue that in many cases it can be difficult to distinguish if data have been generated by a nonlinear model with a smooth structural break in parameters or a random walk. A non-stationary process, rather than a stationary autoregressive process, might in many cases be a more plausible null hypothesis in an LSTAR(1) model.

To obtain our tests, and considerably simplify the testing procedures, we make a Taylor approximation of the smooth transition function yielding an auxiliary model. The inference about unit roots is then based on the OLS estimates from the auxiliary regression model. Analytical expressions for the asymptotic distributions of the tests are enabled by a very applicable Lemma, given and proved in the Mathematical appendix, which generalizes many asymptotic results derived in Phillips (1987), Phillips and Perron (1988), and Hamilton (1994) among others. The asymptotic results hold under a wide class of error terms, and nuisance parameter free tests are presented. All tests are invariant with respect to a possible drift in the random walk.

Despite the fact that the asymptotic results are robust against a wide class of errors, the tests suffer from rather large size distortions, as many of the conventional unit root tests do, in finite-samples. The size distortions are especially severe in the presence of a moving average error structure.

The power of our test, T_a, based on a first-order approximation is compared to the power of the Dickey and Fuller (DF) and Phillips and Perron (PP) tests, when the DGP is an LSTAR(1) model. We conclude that, in gen-

eral, the DF test is the best test when a nearly linear LSTAR(1) model is considered. Studying LSTAR(1) models with more pronounced nonlinearities, both the T_a and PP tests perform better than the DF test. The main reason for this is that the DF test is based on a linear model without a time trend. Moreover, under modest nonlinear changes in the intercept and in the dynamics in the LSTAR(1) model, the power for the PP test is higher than for our T_a test. However, with a modest change in the intercept and a more pronounced nonlinear change in the dynamics, the T_a test performs better than the PP test. It is also revealed that with a dynamic root that is close to unity at the end of the sample period the impact of the level leverage effects is large and both the T_a and PP tests have modest power. As such, we also introduced a parameter constancy test based on a third-order Taylor approximation which in contrary to the T_a and PP tests performs very satisfactorily in the cases of clear nonlinear changes both in the intercept and in the dynamics.

Appendix A

Mathematical Appendix

To be able to prove Theorem 1.1, we first introduce the following lemma.

Lemma 1 *If $\{u_t\}_{t=1}^{\infty}$ satisfies Assumption 1 and $\xi_t = \xi_{t-1} + u_t$, with $P(\xi_0 = 0) = 1$, then as $T \to \infty$*

$$T^{-(p+q/2+1)} \sum_{t=1}^{T} t^p \xi_{t-1}^q \xrightarrow{d} \sigma^q \int_0^1 r^p W(r)^q dr, \tag{A.1}$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p u_t^2 \xrightarrow{a.s.} \bar{\sigma}_u^2/(p+1), \tag{A.2}$$

$$T^{-(v+1/2)} \sum_{t=1}^{T} t^v u_t \xrightarrow{d} \sigma W(1) - v\sigma \int_0^1 r^{v-1} W(r) dr, \tag{A.3}$$

$$\begin{aligned} &T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t \\ &\xrightarrow{d} \tfrac{1}{2} \left(\sigma^2 W(1)^2 - \sigma^2 p \int_0^1 r^{p-1} W(r)^2 dr - \frac{\bar{\sigma}_u^2}{(p+1)} \right), \end{aligned} \tag{A.4}$$

where $p, v \geq 0$, $q \geq 1$, and $\xrightarrow{a.s.}$ denotes convergence almost surely.

Note that Lemma 1 gives more general results than needed to prove Theorem 1.1. In fact, it enables us to derive the limiting distribution for a parameter constancy test in a first order LSTAR model based on any order of approximation of the logistic transition function with an arbitrary number of transitions.

Proof of (A.1). Define the following cadlag function on $D[0,1]$,[1]

$$W_t(r,\omega) = \frac{1}{\sigma\sqrt{T}}\xi_{[Tr]} = \frac{1}{\sigma\sqrt{T}}\xi_{t-1}, \qquad r \in [\tfrac{t-1}{T}, \tfrac{t}{T}), \qquad t = 1, ..., T,$$

where [.] denotes the integer part of its argument. Using $([Tr]+1)/T = t/T$, we can conclude that

$$(([Tr]+1)/T)^p\, W_t(r)^q = (t/T)^p \left(\frac{1}{\sigma\sqrt{T}}\xi_{t-1}\right)^q$$

holds. The left-hand side now defines a continuous functional of $W_t(r)$ on $D[0,1]$. It follows from the Functional Central Limit Theorem (FCLT), the Continuous Mapping Theorem (CMT), and $\lim_{T\to\infty}(([Tr]+1)/T)^p = r^p$, that

$$\begin{aligned} & T^{-(p+q/2+1)} \sum_{t=1}^{T} t^p \xi_{t-1}^q \\ = \; & \sigma^q T^{-1} \sum_{t=1}^{T} (t/T)^p \left(\frac{1}{\sqrt{T}\sigma}\xi_{t-1}\right)^q \\ = \; & \sigma^q \int_0^1 (([Tr]+1)/T)^p\, W_t(r)^q dr \xrightarrow{d} \sigma^q \int_0^1 r^p W(r)^q dr. \end{aligned}$$

Thus, (A.1) holds. ■

Proof of (A.2). Note that the left-hand side in (A.2) can be written as

$$T^{-(p+1)} \sum_{t=1}^{T} t^p u_t^2 = T^{-1} \sum_{t=1}^{T} (t/T)^p \left(u_t^2 - \sigma_t^2\right) + T^{-1} \sum_{t=1}^{T} (t/T)^p \sigma_t^2,$$

where $\sigma_t^2 = E(u_t^2)$. The condition $\sup_{t\in\mathbf{N}} E\left|u_t\right|^\beta < \infty$, where $\beta > 2$ and $(t/T)^p \in [0,1]$ implies that $\sum_{t=1}^{\infty}\left(E(t/T)^p \left|u_t^2 - \sigma_t^2\right|^\beta / t^\beta\right)^{1/\beta} < \infty$ holds. It follows by the strong law of large numbers for α -mixings (see, e.g. McLeish (1975), Theorem 2.10, and Herrndorf (1984)), that $T^{-1}\sum_{t=1}^{T}(t/T)^p\left(u_t^2 - \sigma_t^2\right) \xrightarrow{a.s.} 0$. The second term on the right-hand side converges to $\bar{\sigma}_u^2/(p+1)$ where $\bar{\sigma}_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sigma_t^2$. ■

Proof of (A.3). First, the case where $v = 0$ follows by Wooldridge and White (1988) which says that $T^{-1/2}\sum_{t=1}^{T} u_t \xrightarrow{d} \sigma W(1) \sim N(0,\sigma^2)$. Furthermore, for some integer $v \geq 1$ we use the result in (A.1) letting $p = v-1$ and

[1] See for instance Billingsley (1968) and Davidson (1994).

$q = 1$, to deduce that

$$\begin{aligned}
& T^{-(v+1/2)} \sum_{t=1}^{T} t^{v-1} \xi_{t-1} \\
= \; & T^{-(v+1/2)} \sum_{t=1}^{T} \left(\sum_{i=1}^{T} i^{v-1} - \sum_{j=1}^{t} j^{v-1} \right) u_t \\
= \; & T^{-(v+1/2)} \sum_{t=1}^{T} u_t \sum_{i=1}^{T} i^{v-1} - T^{-(v+1/2)} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} j^{v-1} \right) u_t.
\end{aligned}$$

Rewriting $\sum_{i=1}^{T} i^{v-1} = T^v/v + \mathcal{O}(T^{v-1})$ and $\sum_{j=1}^{t} j^{v-1} = t^v/v + \mathcal{O}(t^{v-1})$, gives

$$\begin{aligned}
& T^{-(v+1/2)} \sum_{t=1}^{T} u_t t^v \\
= \; & T^{-1/2} \sum_{t=1}^{T} u_t - T^{-(v+1/2)} v \sum_{t=1}^{T} t^{v-1} \xi_{t-1} + \mathcal{O}_p(T^{-1}) \\
& \xrightarrow{d} \sigma W(1) - \sigma v \int_0^1 r^{v-1} \; W(r) dr,
\end{aligned}$$

and (A.3) is proved. ■

Proof of (A.4). This is a bit more problematic since we cannot define any continuous functional with bounded variation, almost surely, corresponding to the expression in (A.4). We shall solve this problem by two different approaches.

Approach 1: Define a random polygon function on $C[0,1]$,

$$\begin{aligned}
W_t'(r,\omega) &= \frac{1}{\sigma\sqrt{T}} \xi_{[Tr]}(\omega) + \frac{Tr - [Tr]}{\sigma\sqrt{T}} u_{[Tr]+1}(\omega) \\
&= \frac{1}{\sigma\sqrt{T}} \xi_{t-1}(\omega) + \frac{1}{\sigma\sqrt{T}} \frac{u_t(\omega)}{1/T} r - \frac{t-1}{T},
\end{aligned}$$

which has continuous sample paths with bounded variation, almost surely. Note that $W_t'(r,\omega)$ is linear on each of the subintervals $r \in \left[\frac{t-1}{T}, \frac{t}{T}\right)$and taking on the value $\xi_t/\sqrt{T}\sigma$ at the point t/T. Since $W_t'(r,\omega)$ has bounded variation,

almost surely, we can define the Riemann Stieltjes integral

$$
\begin{aligned}
&\int_0^1 r^p W_t'(r) dW_t'(r) \\
= &\sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} r^p \left(\frac{1}{\sigma\sqrt{T}} \xi_{t-1} + \frac{1}{\sigma\sqrt{T}} \frac{u_t}{1/T} \left(r - \frac{t-1}{T} \right) \right) dW_t'(r) \\
= &\sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \left(r^p \left(\frac{\xi_{t-1} - (t-1){u_t}^2}{\sigma^2} \right) + r^{p+1} \frac{{u_t}^2 T}{\sigma^2} \right) dr \\
= &\ T^{-(p+1)} \sum_{t=1}^{T} \left(\frac{t^p \xi_{t-1} u_t}{\sigma^2} + \frac{t^p {u_t}^2}{2\sigma^2} \right) + \mathcal{O}_p(T^{-1}), \qquad \text{(A.5)}
\end{aligned}
$$

where we at the second equality used that $dW_t'(r) = (\sqrt{T} u_t/\sigma) dr$, to obtain the result, by direct integration, at the last line. Moreover, because the integral in (A.5) is defined in a Riemann-Stieltjes sense, we use the rules for partial integration to obtain the following result for the left-hand side in (A.5)

$$
\int_0^1 r^p W_t' dW_t'(r) = 0.5 \left(W_t'(1)^2 - p \int_0^1 r^{p-1} W_t'(r)^2 dr \right). \qquad \text{(A.6)}
$$

Combining (A.5) and (A.6), making use of the FCLT, CMT, and the result in (A.2), we deduce that

$$
\begin{aligned}
&T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t \\
= &\ \sigma^2/2 \left(W_t'(1)^2 - p \int_0^1 r^{p-1} W_t'^2(r) dr \right) - T^{-(p+1)}/2 \sum_{t=1}^{T} t^p {u_t}^2 + \mathcal{O}_p(T^{-1}) \\
&\xrightarrow{d} 0.5 \left(\sigma^2 W(1)^2 - \sigma^2 p \int_0^1 W(r)^2 r^{p-1} dr \ - \bar{\sigma}_u^2/(p+1) \right),
\end{aligned}
$$

and (A.4) follows.

Approach 2: Above result could also be established without using the theory of Riemann-Stieltjes integrals. To see this note that

$$
\begin{aligned}
& T^{-(p+1)} \sum_{t=1}^{T} t^{p-1} \xi_{t-1}^2 \\
= \; & T^{-(p+1)} \sum_{t=1}^{T} \left(\sum_{i=1}^{T} i^{p-1} - \sum_{j=1}^{t} j^{p-1} \right) (u_t^2 + 2u_t \xi_{t-1}) \\
= \; & T^{-(p+1)} \sum_{t=1}^{T} (u_t^2 + 2u_t \xi_{t-1}) \sum_{i=1}^{T} i^{v-1} - T^{-(p+1)} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} j^{v-1} \right) (u_t^2 + 2u_t \xi_{t-1}).
\end{aligned}
$$

Rewriting $\sum_{i=1}^{T} i^{p-1} = T^p/p + \mathcal{O}(T^{p-1})$ and $\sum_{j=1}^{t} j^{v-1} = t^p/p + \mathcal{O}(t^{p-1})$ gives

$$
\begin{aligned}
& T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t \\
= \; & T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_t + 0.5 T^{-1} \sum_{t=1}^{T} u_t^2 - 0.5 T^{-(p+1)} p \sum_{t=1}^{T} t^{p-1} \xi_{t-1}^2 \\
& -0.5 T^{-(p+1)} \sum_{t=1}^{T} t^p u_t^2 + \mathcal{O}_p(T^{-1}) \\
= \; & 0.5 \sigma^2 W_t(1)^2 - 0.5 T^{-(p+1)} p \sum_{t=1}^{T} t^{p-1} \xi_{t-1}^2 \\
& -0.5 T^{-(p+1)} \sum_{t=1}^{T} t^p u_t^2 + \mathcal{O}_p(T^{-1}) \\
& \xrightarrow{d} 0.5 \left(\sigma^2 W(1)^2 - \sigma^2 p \int_0^1 W(r)^2 r^{p-1} dr - \bar{\sigma}_u^2/(p+1) \right),
\end{aligned}
$$

where we have used $T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_t = 0.5 \sigma^2 W_t(1)^2 - 0.5 T^{-1} \sum_{t=1}^{T} u_t^2$ to obtain the last equality, and thereafter the weak convergence follows by using the results derived in (A.1) and (A.2). The claim in (A.4) follows once again. ■

Proof of Theorem 1.1. The limiting distribution is obtained by applying the results in Lemma 1. To see this note that,

$$
\gamma_T(\hat{\psi} - \psi) = \begin{bmatrix} \hat{\mathbf{M}}_1 & \hat{\mathbf{M}}_2 \\ \hat{\mathbf{M}}_2' & \hat{\mathbf{M}}_3 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{P}}_1 \\ \hat{\mathbf{P}}_2 \end{bmatrix}, \tag{A.7}
$$

where

$$\begin{aligned}
\hat{\mathbf{M}}_1 &= [\hat{m}_{1ij}]_{3\times 3}\,, \text{ and } \hat{m}_{1ij} = T^{-(i+j-1)} \sum\nolimits_{t=1}^{T} t^{(i+j-2)}, \\
\hat{\mathbf{M}}_2 &= [\hat{m}_{2ij}]_{3\times 2}\,, \text{ and } \hat{m}_{2ij} = T^{-(i+j-1/2)} \sum_{t=1}^{T} t^{(i+j-2)} \xi_{t-1}, \\
\hat{\mathbf{M}}_3 &= [\hat{m}_{3ij}]_{2\times 2}\,, \text{ and } \hat{m}_{3ij} = T^{-(i+j)} \sum_{t=1}^{T} t^{(i+j-2)} \xi_{t-1}^2, \\
\hat{\mathbf{P}}_1 &= [\hat{p}_{1i}]_{3\times 1}\,, \text{ and } \hat{p}_{1i} = T^{-(i-1/2)} \sum_{t=1}^{T} t^{i-1} u_t, \\
\hat{\mathbf{P}}_2 &= [\hat{p}_{2i}]_{2\times 1}\,, \text{ and } \hat{p}_{2i} = T^{-i} \sum_{t=1}^{T} t^{i-1} \xi_{t-1} u_t.
\end{aligned}$$

It follows from Lemma 1 that

$$\begin{aligned}
&\hat{\mathbf{M}}_1 \xrightarrow{d} \mathbf{M}_1, \quad \hat{\mathbf{M}}_2 \xrightarrow{d} \sigma \mathbf{M}_2, \quad \hat{\mathbf{M}}_3 \xrightarrow{d} \sigma^2 \mathbf{M}_3, \\
&\hat{\mathbf{P}}_1 \xrightarrow{d} \sigma \mathbf{\Pi}, \quad \hat{\mathbf{P}}_2 \xrightarrow{d} 0.5\sigma^2 \mathbf{\Pi}_2,
\end{aligned}$$

hold, where $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$, $\mathbf{\Pi}_1$, and $\mathbf{\Pi}_2$, are defined in Theorem 1.1. Because (A.7) defines a continuous function we conclude that (1.11) holds.

The consistency of the OLS estimators is an immediate consequence of (A.7). ■

Proof of (1.13) in Corollary 1.2. Define the matrices

$$\begin{aligned}
\mathbf{D}_1 &= diag\left\{ \begin{array}{ccccc} 1 & 1 & 1 & \sigma & \sigma \end{array} \right\}, \\
\tilde{\mathbf{\Pi}}_{12} &= \begin{bmatrix} 0.5(W(1)^2 - 1) \\ 0.5(W(1)^2 - \int W^2 - 0.5) \end{bmatrix}.
\end{aligned}$$

We can now decompose $\mathbf{\Psi}^{-1}\mathbf{\Pi}$ as

$$\mathbf{\Psi}^{-1}\mathbf{\Pi} = \sigma \mathbf{D}_1^{-1} \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{\Pi}_1 \\ \tilde{\mathbf{\Pi}}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \vartheta^2/2\sigma^2 \\ \vartheta^2/4\sigma^2 \end{bmatrix} \right),$$

where $\vartheta^2 = \sigma^2 - \bar{\sigma}_u^2$ and $\mathbf{0}$ is a (3×1) vector. It follows from Theorem 1.1 that $T_n = \mathbf{r}_1 \gamma_T(\hat{\psi} - \psi) \xrightarrow{d} \mathbf{r}_1 \mathbf{\Psi}^{-1}\mathbf{\Pi}$, and where we can write

$$\mathbf{r}_1 \mathbf{\Psi}^{-1}\mathbf{\Pi} = \mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Pi}_1 \\ \tilde{\mathbf{\Pi}}_2 \end{bmatrix} + \frac{\vartheta^2}{\sigma^2} \mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ 1/2 \\ 1/4 \end{bmatrix}, \tag{A.8}$$

because $\mathbf{r}_1\sigma\mathbf{D}_1^{-1} = \mathbf{r}_1$ holds. We see that (A.8) involves the inversion of a 5×5 matrix, but the pre-multiplication with $\mathbf{r}_1$ implies that it is only the elements at the 4th row of the inverse that are of interest. Hence, to proceed we use the rules for the inverse of partioned matrices

$$\begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{M}_1^{-1} + \mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1} & -\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{H}^{-1} \\ -\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1} & \mathbf{H}^{-1} \end{pmatrix}, \tag{A.9}$$

where $\mathbf{H} = \mathbf{M}_3 - \mathbf{M}_2'\mathbf{M}_1^{-1}\mathbf{M}_2$. Obviously we must find the expressions for $-\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1}$ and $\mathbf{H}^{-1}$. Thus, we adopt the following notation

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix},$$

with elements

$$\begin{aligned}
H_{11} &= \int W^2 + 72\int W\int rW - 60\int W\int r^2W + 360\int rW\int r^2W \\
&\quad -9\left(\int W\right)^2 - 192\left(\int rW\right)^2 - 180\left(\int r^2W\right)^2, \\
H_{12} &= \int rW^2 - 9\int W\int rW + 36\int W\int r^2W - 30\int W\int r^3W \\
&\quad -222\int rW\int r^2W + 180\int rW\int r^3W - 180\int r^2W\int r^3W \\
&\quad +36\left(\int rW\right)^2 + 180\left(\int r^2W\right)^2, \\
H_{21} &= H_{12}, \\
H_{22} &= \int r^2W^2 + 72\int rW\int r^2W - 60\int rW\int r^3W + 360\int r^2W\int r^3W \\
&\quad -9\left(\int rW\right)^2 - 192\left(\int r^2W\right)^2 - 180\left(\int r^3W\right)^2,
\end{aligned}$$

where for instance $\int W^2$ is short for $\int_0^1 W(r)^2dr$. This implies that

$$\mathbf{H}^{-1} = \frac{1}{D}\begin{bmatrix} H_{22} & -H_{12} \\ -H_{12} & H_{11} \end{bmatrix} = \begin{bmatrix} H_{44}^* & -H_{45}^* \\ -H_{54}^* & H_{55}^* \end{bmatrix},$$

where $D = H_{11}H_{22} - H_{12}^2$, $H_{44}^* = D^{-1}H_{22}$, $H_{45}^* = H_{54}^* = D^{-1}H_{12}$, and $H_{55}^* = D^{-1}H_{11}$. The sub-indices of the elements with asteriks denotes the actual position they have in (A.9). Due to $\mathbf{r}_1$ we are only interested in H_{44}^*

and H_{45}^*. For the same reason considering the 2×3 matrix $-\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1}$, we only need the result of the upper row (corresponding to the 4th row in (A.9)), denoted $(4,.)$. Hence, $(-\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1})_{(4,.)} = \left[\begin{array}{ccc} H_{41}^* & H_{42}^* & H_{43}^* \end{array}\right]$ where

$$\begin{aligned}
H_{41}^* &= H_{44}^*\left(9\int W - 36\int rW + 30\int r^2W\right) \\
&\quad + H_{45}^*\left(-9\int rW + 36\int r^2W + 30\int r^3W\right) \\
H_{42}^* &= H_{44}^*\left(-36\int W + 192\int rW - 180\int r^2W\right) \\
&\quad + H_{45}^*\left(36\int rW - 192\int r^2W + 180\int r^3W\right) \\
H_{43}^* &= H_{44}^*\left(30\int W - 180\int rW + 180\int r^2W\right) \\
&\quad + H_{45}^*\left(-30\int rW + 180\int r^2W - 180\int r^3W\right).
\end{aligned}$$

This implies that the first term on the right-hand side of (A.8) is given by

$$\begin{aligned}
& \mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Pi}_1 \\ \tilde{\mathbf{\Pi}}_2 \end{bmatrix} \\
= \; & H_{41}^* W(1) + H_{42}^*\left(W(1) - \int W\right) + H_{43}^*\left(W(1) - 2\int rW\right) \\
& + H_{44}^*\left(0.5(W(1)^2 - 1)\right) - H_{45}^*\left(0.5(W(1)^2 - \int W^2 - 0.5)\right) \\
\equiv \; & Q_1(W(r)),
\end{aligned}$$

and that the second term on the right-hand side of (A.8) equals

$$\frac{\vartheta^2}{\sigma^2}\mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ 0.5 \\ 0.25 \end{bmatrix} = \frac{\vartheta^2}{2\sigma^2}\left(H_{44}^* - 0.5H_{45}^*\right) \equiv Q_2(W(r); \sigma^2, \bar{\sigma}_u^2).$$

Adding up $Q_1(W(r))$ and $Q_2(W(r); \sigma^2, \bar{\sigma}_u^2)$ we obtain the limiting distribution for T_n. It is evident that only Q_1 is nuisance parameter free, and the claim in (1.13) follows.

Moreover, to prove (1.14) note first that

$$T^2\Lambda_{44} = s_T^2\mathbf{r}_1\left(\gamma_T\left[\sum_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1}\gamma_T\right)\mathbf{r}_1' \xrightarrow{p} \frac{\bar{\sigma}_u^2}{\sigma^2}\mathbf{r}_1\begin{bmatrix}\mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3\end{bmatrix}^{-1}\mathbf{r}_1' = \frac{\bar{\sigma}_u^2}{\sigma^2}H_{44}^*,$$

$$T^3\Lambda_{45} = s_T^2\mathbf{r}_1\left(\gamma_T\left[\sum_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1}\gamma_T\right)\mathbf{r}_2' \xrightarrow{p} \frac{\bar{\sigma}_u^2}{\sigma^2}\mathbf{r}_1\begin{bmatrix}\mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3\end{bmatrix}^{-1}\mathbf{r}_2' = -\frac{\bar{\sigma}_u^2}{\sigma^2}H_{45}^*,$$

since $s_T^2 \xrightarrow{p} \bar{\sigma}_u^2$ holds under the null hypothesis and that the identities $T\mathbf{r}_1 = \mathbf{r}_1\gamma_T$, and $T^2\mathbf{r}_2 = \mathbf{r}_2\gamma_T$ are valid. We conclude that,

$$\frac{T^2\Lambda_{44}\vartheta^2}{2s_T^2} + \frac{T^3\Lambda_{45}\vartheta^2}{4s_T^2} \xrightarrow{p} \frac{\vartheta^2}{2\sigma^2}H_{44}^* - \frac{\vartheta^2}{4\sigma^2}H_{45}^* = Q_2,$$

implying that

$$T_a = T_n - \left(\frac{T^2\Lambda_{44}\sigma^*}{2s_T^2} + \frac{T^3\Lambda_{45}\sigma^*}{4s_T^2}\right) \xrightarrow{d} Q_1.$$

Thus, (1.14) holds. ■

Proof of Corollary 1.3. Define $\tilde{\gamma}_T = diag\left\{\ T^{1/2}\quad T^{5/2}\quad T\quad T^2\ \right\}$. Note that $\tilde{\gamma}_T\mathbf{R} = \mathbf{R}\gamma_T$ holds. It follows that

$$\begin{aligned}
F_{OLS} &\equiv \left(\mathbf{R}\hat{\psi} - \mathbf{r}\right)'\left\{s_T^2\mathbf{R}\left[\sum_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1}\mathbf{R}'\right\}^{-1}\left(\mathbf{R}\hat{\psi} - \mathbf{r}\right)/4 \\
&= \left(\hat{\psi} - \psi\right)'\mathbf{R}'\tilde{\gamma}_T\left\{s_T^2\tilde{\gamma}_T\mathbf{R}\left[\sum_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1}\mathbf{R}'\tilde{\gamma}_T\right\}^{-1}\tilde{\gamma}_T\mathbf{R}\left(\hat{\psi} - \psi\right)/4 \\
&= \left(\mathbf{R}\gamma_T\left(\hat{\psi} - \psi\right)\right)'\left\{s_T^2\mathbf{R}\gamma_T\left[\sum_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1}\gamma_T\mathbf{R}'\right\}^{-1}\left(\mathbf{R}\gamma_T\left(\hat{\psi} - \psi\right)\right)/4.
\end{aligned}$$

Moreover, $\gamma_T(\hat{\psi} - \psi) \xrightarrow{d} \mathbf{\Psi}^{-1}\mathbf{\Pi}$ holds by Theorem 1.1, and s_T^2 is an consistent estimate to $\bar{\sigma}_u^2$, so the Slutsky Theorem yields

$$F_{OLS} \xrightarrow{d} \left(\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{\Pi}\right)'\left\{\bar{\sigma}_u^2\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{R}'\right\}^{-1}\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{\Pi}/4,$$

and thus, (1.16) holds.

Furthermore, assuming that $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence implies that $\bar{\sigma}_u^2 = \sigma^2$. Define $\mathbf{D}_2 = diag\left\{ \begin{array}{cccc} \sigma & \sigma & 1 & 1 \end{array} \right\}$. We obtain

$$\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{\Pi} = \mathbf{D_2}\mathbf{R}\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{\Pi}_1 \\ \mathbf{\Pi}_2/\mathbf{2} \end{bmatrix}, \tag{A.10}$$

and

$$\sigma^2\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{R}' = \mathbf{D}_2\mathbf{R}\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1}\mathbf{R}'\mathbf{D}_2. \tag{A.11}$$

Using (A.10) and (A.11) yields

$$\begin{aligned} & F_{OLS} \xrightarrow{d} \left(\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{\Pi}\right)'\left\{\sigma^2\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{R}'\right\}^{-1}\mathbf{R}\mathbf{\Psi}^{-1}\mathbf{\Pi}/4 \\ = & \begin{bmatrix} \mathbf{\Pi}_1 \\ \mathbf{\Pi}_2/2 \end{bmatrix}'\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \times \mathbf{R}'\left\{\mathbf{R}\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1}\mathbf{R}'\right\} \\ & \times\mathbf{R}\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{\Pi}_1 \\ \mathbf{\Pi}_2/2 \end{bmatrix}/4. \end{aligned}$$

Finally, since $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence, $\mathbf{\Pi}_2$ is nuisance parameter free. Thus, (1.17) holds. ■

Bibliography

ANDREWS, D. W. (1993): "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 4, 821–856.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures.* John Wiley.

CAMPBELL, J. Y., AND P. PERRON (1991): "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," *NBER Macroeconomics Annual*, pp. 141–200.

CHU, C. S., AND H. WHITE (1991): "Testing for structural change in some simple time series models," University of California, San Diego, CA, Discussion paper no. 91-6.

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

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series," *Journal of the American Statistical Association*, 74, 427–431.

HALL, A. (1989): "Testing for a unit root in the presence of moving average errors," *Biometrika*, 76, 49–56.

HAMILTON, J. D. (1994): *Time Series Analysis.* Princeton University Press, New Jersey.

HANSEN, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

HERRNDORF, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12, 141–153.

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

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

MCLEISH, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 3, 829–839.

NYBLOM, J. (1989): "Testing for the constancy of parameters over time," *Journal of the American Statistical Association*, 84, 223–230.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

SCHWERT, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–159.

SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990): "Inference in Linear Time Series with Some Unit Roots," *Econometrica*, 58, 113–144.

STOCK, J. H., AND M. W. WATSON (1996): "Evidence on structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

WOOLDRIDGE, J. M., AND H. WHITE (1988): "Some invariance principles and central limit theorems for dependent heterogeneous processes," *Econometric Theory*, 4, 210–230.

Chapter 2

Dickey-Fuller Type of Tests Against Nonlinear Dynamic Models

2.1 Introduction

It is well known that classical unit root tests based on linear models such as those by Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988) among others, lack power when the model specification under the alternative hypothesis is nonlinear. In particular, nonlinear models with structural changes in levels and trends bias the classical tests towards nonrejection, as pointed out in Perron (1990). In the light of that many time series exhibit jumps or more smooth structural changes in levels and trends over time, the bias in the classical tests is particularly a non-desirable property. As such, this has serious implications for applied work because shocks will be treated as if they have a permanent effect, when they in fact are transitory.

Nowadays there are many ways of how to test the unit root hypothesis in a nonlinear set-up. One approach is to test unit roots in time series models with a change in levels and/or trends where the break point is known as in Perron (1989), Perron (1990), and Lanne and Lütkepohl (2002) among others, or unknown as in Perron and Vogelsang (1992), Zivot and Andrews (1992) and Banerjee, Lumsdaine, and Stock (1992) as examples.

Most of the approaches in above quoted literature deal with a single break in threshold type of models. In many cases this can be seen a too restrictive and models with more smooth and multiple breaks in level and/or trends are called for. Example of more general models can be found in Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), where the changes are determined by a logistic smooth transition function, or as in Saikkonen and Lütkepohl (2002) and Lanne, Lütkepohl, and Saikkonen (2003) where the functional form of the changes is set very flexible.

The approach that we are taking is similar to the one in Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), i.e. the null hypothesis of a unit root is tested against a model with a logistic smooth transition in the intercept and in the time trend. We extend, however, their discussion by explicitly allowing for a smooth transition in the dynamics. This extension is reasoned by e.g. that the amplitude of the fluctuations after a trend shift may not be the same as before the trend shift. As such, the models that we consider are the pth order logistic smooth transition autoregressive model (LSTAR(p)) in Lin and Teräsvirta (1994), and also an LSTAR(p) model where a time trend is included (LSTART(p)). We assume that the transitions take place over time and that the speed of transition between regimes may not be the same for all parameters.

We derive several unit root tests in the aforementioned LSTAR type of models and analytical limiting distributions for the tests are presented. This

is in contrary to Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002) because the limiting distributions of their unit root tests are found by simulations. Their testing methodology is also different and is based on NLS estimation, whereas our tests are based on a Taylor approximation of the transition function which implies that inference for unit roots is easily obtained by simple regressions.

It should be mentioned that there already exists several unit root tests in the smooth transition autoregressive (STAR) framework, see for instance Enders and Granger (1998), Bec, Salem, and Carrasco (2002), Eklund (2003a), Eklund (2003b), and Kapetanios, Shin, and Snell (2003), but they differ, however fundamentally since they do *not* model the transition as a function of time, it is modelled as a function of lagged dependent variables.

The rest of the chapter is organized as follows. In Section 2.2 the models are presented. The procedure for testing a unit root against these nonlinear models is described in Section 2.3. In Section 2.4 theoretical asymptotic properties of the tests are presented. Finite-sample properties of the tests are investigated in Section 2.5. Concluding remarks are given in Section 2.6. Thereafter two appendices follow where proofs and additional tables can be found.

2.2 The models

In this section we introduce a new family of nonlinear models by adding a time trend with its own individual transition function to the LSTAR model defined in Lin and Teräsvirta (1994). Such a model could be defined as

$$y_t = \mathbf{x}_t'\boldsymbol{\pi}_1 + \mathbf{x}_t'\mathbf{F}\boldsymbol{\pi}_2 + u_t, \quad t = 1, ..., T, \tag{2.1}$$

where $\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, ..., y_{t-p}, t)'$ is a $(p+2) \times 1$ vector and p is an integer such that $p \geq 1$, $\boldsymbol{\pi}_1 = (\pi_{10}, ..., \pi_{1,p+1})' \in \mathbb{R}^{p+2}$, $\mathbf{F} = diag\{F_0, ..., F_{p+1}\}$ is a $(p+2) \times (p+2)$ matrix with transition functions on its diagonal, $\boldsymbol{\pi}_2 = (\pi_{20}, ..., \pi_{2,p+1})' \in \mathbb{R}^{p+2}$, and u_t is an error term whose properties are discussed later on. Suitable transition functions could be defined according to

$$F_i(t; \boldsymbol{\phi}_i) = \frac{1}{1 + \exp\{-\gamma_i(t - c_i)\}} - \frac{1}{2}, \quad i = 0, ..., p+1, \tag{2.2}$$

where $\boldsymbol{\phi}_i = (\gamma_i, c_i) \in \mathbb{R}_{++} \times (0, T)$. Viewing $F_i(t, \boldsymbol{\phi}_i)$ as a function of t, with $\boldsymbol{\phi}_i$ fixed, it is a bounded continuous non-decreasing transition function in t such that $F_i(t) : \mathbb{R}_+ \to [-1/2, 1/2]$. The specification in (2.2) allows for one transition over time for each parameter, where γ_i determines the speed of transitions from one regime to another, and the parameter c_i indicates which

point in time the transition is symmetric about.[1] Viewing $F_i(t;\gamma_i,c_i)$ as a function of γ_i, with t and c_i fixed, $\lim_{\gamma_i\to\infty} F_i(\gamma_i) = -0.5$ if $t \in (c_i, T]$, and $\lim_{\gamma_i\to\infty} F_i = 0.5$ if $t \in [0, c_i]$, meaning that the transition functions make a jump (point of discontinuities) at c_i. On the other hand, $\lim_{\gamma_i\to 0} F_i(\gamma_i) = 0$, and the resulting model in (2.1) is linear. Finally we note that $F_i(t;\gamma_i,c_i) \in C^n(\mathbb{R}_+, \mathbb{R}_{++} \times (0,T))$ where n is an integer such that $n \geq 1$.[2]

It is evident that the model in (2.1) with (2.2) is able to capture the properties of quite complex nonlinear time series where e.g. the level, trend and dynamics are initially at an equilibrium and as time evolves a nonlinear adjustment towards a new long-run equilibrium takes place. Notice also that the model specification in (2.1) with (2.2) nests models such as the LSTAR(p) model introduced in Lin and Teräsvirta (1994), the classical linear autoregressive models in Dickey and Fuller (1979) and Phillips and Perron (1988), many of the nonlinear models in Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), and the threshold autoregressive (TAR) model discussed in Chan and Tong (1986).

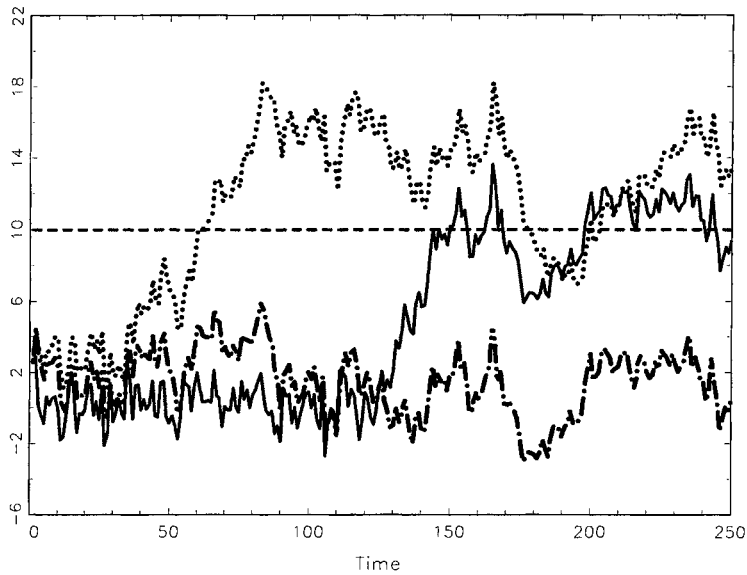
2.2.1 The LSTAR(p), LSTART(p), and LSTD(p) models

The discussion in this chapter will focus, without loss of generality due to our testing methodology in the next section, on a sub-class of models in (2.1) with (2.2), where all transition functions are set equal. The generic transition function will hereafter be denoted F and is defined as F_i in (2.2) but with all indices dropped. Within the sub-class of models characterized by equal transition functions, there are three models of particular interest. First, the LSTAR(p) model obtained by letting $\pi_{1,p+1} = \pi_{2,p+1} \equiv 0$ in (2.1). Second, a new model that we call the LSTART(p) model which is characterized by the same nonlinearities as the LSTAR(p) model but in addition accommodates a nonlinear trend. To obtain the LSTART(p) model, no restrictions in (2.1) are imposed. Finally, another new model with a smooth transition only in the deterministic part, i.e. in the intercept and the time trend, is considered. This model is obtained by letting $\pi_{21} = ... = \pi_{2,p} \equiv 0$ in (2.1), and is referred to as the LSTD(p) model. Typical realizations of an LSTAR and LSTART model together with their linear counterparts are depicted in Figures 2.1 and 2.2, respectively.

[1]The concern in this chapter is a smooth transition function with a single transition between regimes. However, in the Appendix A it becomes clear that we could generalize the discussion to transition functions with an arbitrary number of transitions, i.e. let $F(t;\gamma,c_1,...,c_m) = 1/(1+e^{-\gamma\Pi_{i=1}^{m}(t-c_i)}) - 1/2$, where $\gamma > 0$, and $c_1 < ... < c_m$ for $m \geq 1$.

[2]$C^n(\mathbb{R}_+, \mathbb{R}_{++} \times (0,T))$ denotes the family of continuous functions that are differentiable of order n with respect to $t \in \mathbb{R}_+$, $\gamma_i \in \mathbb{R}_{++}$, and $c_i \in (0,T)$.

Figure 2.1 Typical realizations of an LSTAR(1) model (solid line) with a long-run equilibrium (dashed line), a stationary AR(1) process (dash-dotted line), and a random walk (dotted line).

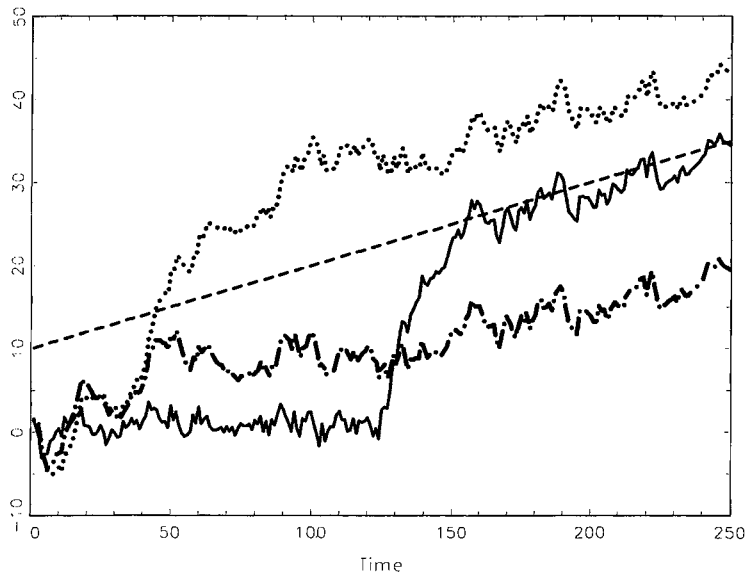


In the coming discussion it is shown that the power of conventional unit root tests such as those by Dickey and Fuller (1979) and Phillips and Perron (1988) have power close to zero under the nonlinear models illustrated in Figures 2.1 and 2.2. In these figures it is also seen that a random walk is a relevant null hypothesis in LSTAR type of models, i.e. the trajectories of a random walk in small samples may appear similarly to the trajectories generated by an LSTAR type of model.

2.3 Testing methodology

In this section we derive several testing equations that serve as devices for the inference about unit roots in the LSTAR, LSTART and LSTD models proposed above. To proceed it is convenient to separate the cases $p = 1$ and $p \geq 2$. When $p = 1$, the testing equations are called nonlinear Dickey-Fuller (NDF) testing equations. Letting $p \geq 2$, the resulting testing equations are abbreviated to the nonlinear augmented Dickey-Fuller (NADF) testing

Figure 2.2 Typical realizations of an LSTART(1) model (solid line) with a long-run equilibrium (dashed line), a trend stationary AR(1) process (dash-dotted line), and a random walk with drift (dotted line).



equations, as well as nonlinear partially augmented Dickey-Fuller (NPADF) testing equations.

2.3.1 The NDF testing equations

Assuming that $p = 1$ we want to test the null hypothesis of a random walk without drift against the LSTAR(1) model. This is reasoned by that both models lack the property of a time trend. Furthermore, we also want to test the null hypothesis of a random walk with drift against the LSTART(1) and the LSTD(1) models. As such, both the null and the alternative hypotheses display a trend, but are fundamentally different since in the former case the trend is stochastic and in the latter case it is nonlinear and deterministic. Testing the unit root hypothesis in the LSTAR(1), LSTART(1), and LSTD(1) models is formalized as follows

$$\begin{aligned} H_{01} &: \; y_t = y_{t-1} + u_t, \\ H_{a1} &: \; y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1}) F(t, \phi) + u_t, \end{aligned}$$

$$
\begin{aligned}
H_{0m} &: \; y_t = \pi_{10} + y_{t-1} + u_t, \quad m = 2, 3, \\
H_{a2} &: \; y_t = \pi_{10} + \pi_{11} y_{t-1} + \pi_{12} t + (\pi_{20} + \pi_{21} y_{t-1} + \pi_{22} t) F(t, \phi) + u_t, \\
H_{a3} &: \; y_t = \pi_{10} + \pi_{11} y_{t-1} + \pi_{12} t + (\pi_{20} + \pi_{22} t) F(t, \phi) + u_t.
\end{aligned}
$$

The implementation of these three tests is straightforward in the sense that for $m = 1, 2, 3$ the models under the alternative hypotheses, H_{am}, nest the models under the null hypotheses, H_{0m}. It is evident that these tests can be conducted by imposing the parameter restrictions

$$\gamma = 0, \quad \pi_{10} = 0, \quad \pi_{11} = 1, \tag{2.3}$$

on the LSTAR(1) model represented by H_{a1} to obtain the null hypothesis H_{01}, and imposing the restrictions

$$\gamma = 0, \quad \pi_{10} \in \mathbb{R}, \quad \pi_{11} = 1, \quad \pi_{12} = 0, \tag{2.4}$$

on the LSTART(1) and the LSTD(1) models represented by H_{a2} and H_{a3}, respectively, to deduce the null hypotheses H_{02} and H_{03}. However, it is known that letting $\gamma = 0$ leads to an identification problem under the null hypotheses. Following Luukkonen, Saikkonen, and Teräsvirta (1988) we remedy this problem by a third-order Taylor expansion of γ around 0 in F. This is feasible because $F \in C^4[\mathbb{R}_+, \mathbb{R}_{++} \times (0, T)]$, and since F is an odd function we especially note that $\partial F / \partial \gamma|_{\gamma=0} \neq 0$ and $\partial^3 F / \partial \gamma^3|_{\gamma=0} \neq 0$ in an open interval $(-\varepsilon, \varepsilon)$, for $\varepsilon > 0$. The motivation of a third-order Taylor approximation, rather than a first, is mainly the robustness in power to a wider class of DGP's under the alternative hypotheses. As far as the LSTAR model is concerned, it is well known that a first-order Taylor approximation results in low power if the transition only takes place in the intercept, see Luukkonen, Saikkonen, and Teräsvirta (1988) and the discussion in Chapter 1. Similar arguments hold for the LSTART and LSTD models and a third-order approximation is in many cases strongly preferable. Thus, a third-order Taylor approximation of the transition function is given by

$$T_3(t; \gamma, c) = \gamma(t - c)/4 + \gamma^3 (t - c)^3 / 48 + r(\gamma), \tag{2.5}$$

where $r(\gamma)$ is a remainder such that $r(0) = 0$. Substituting (2.5) into the models in H_{a1}, H_{a2}, and H_{a3}, and merging terms yields the auxiliary regression equations[3]

[3]Relaxing the assumption that all transitions are set equal in the transition function (2.2) will not change the set of explanatory variables in H_{a1}^{aux}, H_{a2}^{aux}, and H_{a3}^{aux}. However, the induced parameters in the auxiliary regression will not be the same in terms of the original parameters in the models in H_{a1}, H_{a2}, and H_{a3}.

$$\begin{aligned} H_{a1}^{aux} &: \; y_t = \mathbf{s}_{1t}'\boldsymbol{\lambda}_1 + (y_{t-1}\mathbf{s}_{1t})'\boldsymbol{\varphi}_1 + u_t^*, \\ H_{a2}^{aux} &: \; y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_2 + (y_{t-1}\mathbf{s}_{1t})'\boldsymbol{\varphi}_2 + u_t^*, \\ H_{a3}^{aux} &: \; y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_3 + y_{t-1}\varphi_{30} + u_t^*, \end{aligned} \tag{2.6}$$

where $\mathbf{s}_{1t} = (1, t, t^2, t^3)'$, $\boldsymbol{\lambda}_1 = (\lambda_{10}, ..., \lambda_{13})'$, $\boldsymbol{\varphi}_1 = (\varphi_{10}, ..., \varphi_{13})'$, u_t^* is an error term adjusted with respect to the Taylor expansions such that $u_t^* = u_t$ holds whenever $\gamma = 0$, $\mathbf{s}_{2t} = (1, t, ..., t^4)'$, $\boldsymbol{\lambda}_2 = (\lambda_{20}, ..., \lambda_{24})'$, $\boldsymbol{\varphi}_2 = (\varphi_{20}, ..., \varphi_{23})'$, and $\boldsymbol{\lambda}_3 = (\lambda_{30}, ..., \lambda_{34})'$. The corresponding auxiliary null hypotheses are given by

$$\begin{aligned} H_{01}^{aux} &: \; \lambda_{1i} = 0 \;\; \forall i, \quad \varphi_{10} = 1, \quad \varphi_{1j} = 0, \quad j \geq 1, \\ H_{02}^{aux} &: \; \lambda_{20} \in \mathbb{R}, \quad \lambda_{2i} = 0, \quad i \geq 1, \quad \varphi_{20} = 1, \quad \varphi_{2j} = 0 \quad j \geq 1, \\ H_{03}^{aux} &: \; \lambda_{30} \in \mathbb{R}, \quad \lambda_{3i} = 0, \quad i \geq 1, \quad \varphi_{30} = 1. \end{aligned} \tag{2.7}$$

It should be pointed out that the auxiliary regression model for the LSTD(1) model given by H_{a3}^{aux} is the same type of regression model as in Ouliaris, Park, and Phillips (1989), but in addition they allow for a trend of arbitrary order. Applying the same order of the time trends, (i.e. trends up to the fourth-order) we note that their null hypothesis would correspond to our null hypothesis H_{03}^{aux} but where the restrictions on $\lambda_{31}, ..., \lambda_{33}$ are relaxed. In this context the null hypothesis is a random walk where the drift is characterized by a third order polynomial in t, and with implications that are to be stressed later on.

To proceed, the auxiliary null hypotheses H_{02}^{aux} and H_{03}^{aux} imply that the data generating processes (DGP's) are random walks with drift. This leads to problems with multicollinearity in large samples since y_{t-1} is then asymptotically equivalent to $t-1$, an explanatory variable already included in the regression equations in H_{a2}^{aux} and H_{a3}^{aux}. We avoid this problem by the transformation $\xi_t \equiv y_t - \lambda_{m0}t$ for $m = 2, 3$ and where $\lambda_{m0}t = E_{H_{0m}^{aux}}[y_t] = \pi_{10}t$, see e.g. Sims, Stock, and Watson (1990), Hamilton (1994), and Chapter 1 for details. These transformations give rise to the new set of auxiliary regression equations

$$\tilde{H}_{a2}^{aux} \;:\; y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_2^* + \left(\xi_{t-1}\mathbf{s}_{1t}\right)'\boldsymbol{\varphi}_2 + u_t^*, \tag{2.8}$$

$$\tilde{H}_{a3}^{aux} \;:\; y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_3^* + \xi_{t-1}\varphi_{30} + u_t^*, \tag{2.9}$$

where $\boldsymbol{\lambda}_2^* = (\lambda_{20}^*, ..., \lambda_{24}^*)'$, $\xi_{t-1} = y_{t-1} - \pi_{10}(t-1)$, and $\boldsymbol{\lambda}_3^* = (\lambda_{30}^*, ..., \lambda_{34}^*)'$.

The null hypotheses H_{a2}^{aux} and H_{a3}^{aux} transform further and yield

$$\tilde{H}_{02}^{aux}: \quad \lambda_{21}^* \in \mathbb{R}, \quad \lambda_{2i}^* = 0, \; i \neq 1, \quad \varphi_{20} = 1, \quad \varphi_{2j} = 0, \; j \geq 1, \tag{2.10}$$

$$\tilde{H}_{03}^{aux}: \quad \lambda_{31}^* \in \mathbb{R}, \quad \lambda_{3i}^* = 0, \; i \neq 1, \quad \varphi_{30} = 1. \tag{2.11}$$

It is now clear that imposing the parameter restrictions given in (2.7), (2.10), and (2.11) on the corresponding auxiliary regression equations (2.6), (2.8) and (2.9), respectively, the DGP's $y_t = y_{t-1} + u_t$ for $m = 1$, and $\xi_t = \xi_{t-1} + u_t$ for $m = 2, 3$, are obtained. In other words, we end up with a random walk without a drift for *all* cases with the same error term as in (2.1). The inference about unit roots in the LSTAR(1), LSTART(1), and LSTD(1) models is now based on the NDF testing equations given in H_{a1}^{aux}, $\tilde{H}_{a2}^{aux}$, and $\tilde{H}_{a3}^{aux}$, respectively.

2.3.2 The NADF testing equations

Assuming that $p \geq 2$ means that the LSTAR(p), LSTART(p) and LSTD(p) models are in focus. In these augmented models we shall test for a single unit root. The desired testing situations is summarized as follows

$$\begin{aligned}
H_{01}^{a} \quad &: \quad (1 - \sum_{i=1}^{p} \pi_{1i} L^i) y_t = u_t, \\
H_{a1}^{a} \quad &: \quad y_t = \pi_{10} + \pi_{11} y_{t-1} + \ldots + \pi_{1,p} y_{t-p} \\
&\quad\; + (\pi_{20} + \pi_{21} y_{t-1} + \ldots + \pi_{2,p} y_{t-p}) F(t, \phi) + u_t, \\
\\
H_{0m}^{a} \quad &: \quad (1 - \sum_{i=1}^{p} \pi_{1i} L^i) y_t = \pi_{10} + u_t, \quad m = 2, 3, \\
H_{a2}^{a} \quad &: \quad y_t = \pi_{10} + \pi_{11} y_{t-1} + \ldots + \pi_{1,p} y_{t-p} + \pi_{1,p+1} t \\
&\quad\; + (\pi_{20} + \pi_{21} y_{t-1} + \ldots + \pi_{2,p} y_{t-p} + \pi_{2,p+1} t) F(t, \phi) + u_t, \\
H_{a3}^{a} \quad &: \quad y_t = \pi_{10} + \pi_{11} y_{t-1} + \ldots + \pi_{1,p} y_{t-p} + \pi_{1,p+1} t \\
&\quad\; + (\pi_{20} + \pi_{2,p+1} t) F(t, \phi) + u_t,
\end{aligned}$$

where L denotes the lag operator. As before, the models under the alternative nest the models under the null hypotheses and the desirable tests could be carried out by imposing the restrictions $\gamma = 0$ and $\pi_{10} = 0$ on H_{a1}^{a}, and $\gamma = 0$ and $\pi_{10} \in \mathbb{R}$ on H_{a2}^{a} and H_{a3}^{a}, respectively. In addition, we assume, for all m, that $\{y_t\}_{t=1}^{\infty}$ contains a single unit root under the null hypothesis meaning that the characteristic polynomial $c(z) = 1 - \sum_{i=1}^{p} \pi_{1i} z^i = 0$ has a single unit root and all other roots lie outside the unit circle. Furthermore, the restriction $\gamma = 0$ implies the same kind of identification problem that was mentioned in

Section 2.3.1, and is circumvented in the same way. In other words, applying (2.5) to the models represented in H_{a1}^{a}, H_{a2}^{a}, and H_{a3}^{a}, respectively, yields the augmented auxiliary regression equations

$$\begin{aligned} H_{a1}^{a,aux} &: \; y_t = \mathbf{s}_{1t}'\boldsymbol{\lambda}_1^a + (\mathbf{s}_{1t} \otimes \mathbf{y}_{-p})' \boldsymbol{\varphi}_1^a + u_t^*, \\ H_{a2}^{a,aux} &: \; y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_2^a + (\mathbf{s}_{1t} \otimes \mathbf{y}_{-p})' \boldsymbol{\varphi}_2^a + u_t^*, \\ H_{a3}^{a,aux} &: \; y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_3^a + \mathbf{y}_{-p}\boldsymbol{\varphi}_3^a + u_t^*, \end{aligned}$$

where $\boldsymbol{\lambda}_1^a = (\lambda_{10}^a, ..., \lambda_{13}^a)'$, $\mathbf{y}_{-p} = (y_{t-1}, ..., y_{t-p})'$ is a $p \times 1$ vector, $\boldsymbol{\varphi}_1^a = (\boldsymbol{\varphi}_{10}^{a\prime}, ..., \boldsymbol{\varphi}_{13}^{a\prime})'$ and $\boldsymbol{\varphi}_{1n}^a = \left(\varphi_{1n1}^a, ..., \varphi_{1np}^a\right)'$ where $n = 0, ..., 3$, u_t^* is an error term adjusted with respect to the Taylor expansions such that $u_t^* = u_t$ holds whenever $\gamma = 0$, $\boldsymbol{\lambda}_2^a = (\lambda_{20}^a, ..., \lambda_{24}^a)'$, $\boldsymbol{\varphi}_2^a = (\boldsymbol{\varphi}_{20}^{a\prime}, ..., \boldsymbol{\varphi}_{23}^{a\prime})'$ and $\boldsymbol{\varphi}_{2n}^a = \left(\varphi_{2n1}^a, ..., \varphi_{2np}^a\right)'$ where $n = 0, ..., 3$, $\boldsymbol{\lambda}_3^a = (\lambda_{30}^a, ..., \lambda_{34}^a)'$, and $\boldsymbol{\varphi}_3^a = \left(\varphi_{301}^a, ..., \varphi_{30p}^a\right)'$.

Letting $\Delta y_t = y_t - y_{t-1}$, more convenient representations of the auxiliary regression equations above are found as follows.

Lemma 1 *(i) Assuming $p \geq 2$, then the models in $H_{a1}^{a,aux}$ and $H_{a2}^{a,aux}$ can be represented by*

$$y_t = \mathbf{s}_{1t}'\boldsymbol{\lambda}_1^a + (y_{t-1}\mathbf{s}_{1t})'\boldsymbol{\rho}_1^a + (\mathbf{s}_{1t} \otimes \Delta\mathbf{y}_t)' \boldsymbol{\zeta}_1^a + u_t^*, \tag{2.12}$$

$$y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_2^a + (y_{t-1}\mathbf{s}_{2t})'\boldsymbol{\rho}_2^a + (\mathbf{s}_{1t} \otimes \Delta\mathbf{y}_t)' \boldsymbol{\zeta}_2^a + u_t^*, \tag{2.13}$$

where $\boldsymbol{\rho}_m^a = (\rho_{m0}^a, ..., \rho_{m3}^a)'$, $\Delta\mathbf{y}_t = (\Delta y_{t-1}, ..., \Delta y_{t-p+1})'$ is a $(p-1) \times 1$ vector, $\boldsymbol{\zeta}_m^a = (\boldsymbol{\zeta}_{m0}^{a\prime}, ..., \boldsymbol{\zeta}_{m3}^{a\prime})'$ with $\boldsymbol{\zeta}_{mn}^a = (\zeta_{mn1}^a, \zeta_{mn2}^a, ..., \zeta_{mn,p-1}^a)'$ where $n = 0, ..., 3$. Furthermore, the parameter vectors $\boldsymbol{\rho}_m^a$ and $\boldsymbol{\zeta}_m^a$ are defined through the parameter vector $\boldsymbol{\varphi}_m^a$ in $H_{a1}^{a,aux}$ and $H_{a2}^{a,aux}$ by

$$\begin{aligned} \rho_{mn}^a &= \sum_{q=1}^{p} \varphi_{mnq}^a, \\ \zeta_{mn1}^a &= -\sum_{q=2}^{p} \varphi_{mnq}^a, \\ \zeta_{mn2}^a &= -\sum_{q=3}^{p} \varphi_{mnq}^a, \\ &\vdots \\ \zeta_{mn,p-1}^a &= -\varphi_{mnp}^a. \end{aligned}$$

(ii) Assuming $p \geq 2$, then the model in $H_{a3}^{a,aux}$ can be represented by

$$y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_3^a + y_{t-1}\rho_{30}^a + \Delta\mathbf{y}_t'\boldsymbol{\zeta}_3^a + u_t^*, \tag{2.14}$$

where $\zeta_3^a = \left(\zeta_{301}^a, \zeta_{302}^a, ..., \zeta_{30,p-1}^a\right)$. *Moreover,* ρ_{30}^a *and the parameter vector* ζ_3^a *are defined through the parameter vector* φ_3^a *in* $H_{a3}^{a,aux}$ *by*

$$\begin{aligned}
\rho_{30}^a &= \sum_{q=1}^{p} \varphi_{30q}^a, \\
\zeta_{301}^a &= -\sum_{q=2}^{p} \varphi_{30q}^a, \\
\zeta_{302}^a &= -\sum_{q=3}^{p} \varphi_{mnq}^a, \\
&\vdots \\
\zeta_{30,p-1}^a &= -\varphi_{mnp}^a.
\end{aligned}$$

Proof. See Appendix A. ■

In Lemma 1 it is seen that the two first terms on the right-hand sides of (2.12)-(2.14) are the same set of explanatory variables as in the case $p = 1$. The augmented auxiliary regression equations are therefore obtained by adding lagged differences of the dependent variable multiplied by time trends up to order three to the auxiliary regression equations in H_{a1}^{aux}, H_{a2}^{aux}, and H_{a3}^{aux}, respectively. That the characteristic polynomial $c(z) = 1 - \sum_{i=1}^{p} \pi_{1i} z^i = 0$ only contains a single unit root corresponds now to $\rho_{m0}^a = 1$ and that the roots of $c(z) = 1 - \sum_{q=1}^{p-1} \zeta_{m0q}^a z^q = 0$ lie outside the unit circle. The null hypotheses can be expressed as

$$\begin{aligned}
H_{01}^{a,aux}: &\quad \lambda_{1i}^a = 0 \ \ \forall i, \quad \rho_{10}^a = 1, \quad \rho_{1j}^a = 0, \quad j \geq 1, \\
&\quad \zeta_{10q}^a \in \mathbb{R}, \quad \zeta_{1nq}^a = 0, \quad n \geq 1, \quad q = 1, ..., p-1, \\
H_{02}^{a,aux}: &\quad \lambda_{20}^a \in \mathbb{R}, \quad \lambda_{2i}^a = 0, \quad i \geq 1, \quad \rho_{20}^a = 1, \quad \rho_{2j}^a = 0, \quad j \geq 1, \\
&\quad \zeta_{20q}^a \in \mathbb{R}, \quad \zeta_{2nq}^a = 0, \quad n \geq 1, \quad q = 1, ..., p-1, \\
H_{03}^{a,aux}: &\quad \lambda_{30}^a \in \mathbb{R}, \quad \lambda_{3i}^a = 0, \quad i \geq 1, \quad \rho_{30}^a = 1, \\
&\quad \zeta_{3nq}^a \in \mathbb{R}, \quad \zeta_{3nq}^a = 0, \quad n \geq 1, \quad q = 1, ..., p-1.
\end{aligned} \tag{2.15}$$

Imposing the restrictions in $H_{01}^{a,aux}$, $H_{02}^{a,aux}$, and $H_{03}^{a,aux}$ on (2.12), 2.13 and (2.14), respectively, yields

$$\begin{aligned}
\left(1 - \sum_{q=1}^{p-1} \zeta_{10q}^a L^q\right) \Delta y_t &= u_t, \\
\left(1 - \sum_{q=1}^{p-1} \zeta_{10q}^a L^q\right) \Delta y_t &= \lambda_{m0} + u_t.
\end{aligned}$$

Because the roots of $c(z) = 1 - \sum_{q=1}^{p-1} \zeta_{m0q}^{a} z^q = 0$ fall outside the unit circle, it follows that

$$\begin{aligned} H_{01}^{a,aux} &: \Delta y_t = v_t, \\ H_{0m}^{a,aux} &: \Delta y_t = \mu_m + v_t, \quad m = 2,3, \end{aligned}$$

where $v_t = \left(1 - \zeta_{m01}^{a} L - \zeta_{m02}^{a} L^2 - .... - \zeta_{m0,p-1}^{a} L^{p-1}\right)^{-1} u_t$ and $\mu_m = \lambda_{m0}^{a}/(1 - \zeta_{m01}^{a} - \zeta_{m02}^{a} - .... - \zeta_{m0,p-1}^{a})$. For $m = 2,3$, the models under $H_{0m}^{a,aux}$ are random walks with drift, and again we encounter a problem with multicollinearity in the auxiliary regression equations (2.13) and (2.14). The problem is circumvented by defining $\xi_t \equiv y_{t-1} - \mu_m t$ and $\Delta\xi_t \equiv \Delta y_t - \mu_m$ where $\mu_m t = E_{H_{0m}^{a,aux}}[y_t]$ and $\mu_m = E_{H_{0m}^{a,aux}}[\Delta y_t]$. These transformations imply that (2.13) and (2.14) can be expressed as

$$\tilde{H}_{a2}^{a,aux} : \quad y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_2^{a*} + (\xi_{t-1}\mathbf{s}_{1t})'\boldsymbol{\rho}_2^{a} + (\Delta\boldsymbol{\xi}_t \otimes \mathbf{s}_{1t})'\boldsymbol{\zeta}_2^{a} + u_t^*, \tag{2.16}$$

$$\tilde{H}_{a3}^{a,aux} : \quad y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_3^{a*} + \xi_{t-1}\rho_{30}^{a} + \Delta\boldsymbol{\xi}_t'\boldsymbol{\zeta}_3^{a} + u_t^*, \tag{2.17}$$

where $\boldsymbol{\lambda}_2^{a*} = (\lambda_{20}^{a*}, ..., \lambda_{24}^{a*})'$, and $\boldsymbol{\lambda}_3^{a*} = (\lambda_{30}^{a*}, ..., \lambda_{34}^{a*})'$. The null hypotheses transform to

$$\begin{aligned} \tilde{H}_{02}^{a,aux} : \quad & \lambda_{21}^{a*} \in \mathbb{R},\ \lambda_{2i}^{a*} = 0,\ i \neq 1,\ \rho_{20}^{a} = 1,\ \rho_{2j}^{a} = 0,\ j \geq 1, \\ & \zeta_{20q}^{a} \in \mathbb{R},\ \zeta_{2nq}^{a} = 0,\ n \geq 1,\ q = 1, ..., p-1, \end{aligned} \tag{2.18}$$

$$\begin{aligned} \tilde{H}_{03}^{a,aux} : \quad & \lambda_{31}^{a*} \in \mathbb{R},\ \lambda_{3i}^{a*} = 0,\ i \neq 0,\ \rho_{30}^{a} = 1, \\ & \zeta_{3nq}^{a} \in \mathbb{R},\ \zeta_{3nq}^{a} = 0,\ n \geq 1,\ q = 1, ..., p-1. \end{aligned} \tag{2.19}$$

Applying the parameter restrictions (2.15), (2.18), and (2.19) to (2.12), (2.16), and (2.17), respectively, a random walk *without* drift for all m is obtained. Specifically, $m = 1$ implies $y_t = y_{t-1} + v_t$, and $m = 2,3$ gives that $\xi_t = \xi_{t-1} + v_t$. The difference between the random walks derived in this section and the random walks in Section 2.3.1 is that $\{v_t\}$ defines a sequence of serially correlated errors, whereas $\{u_t\}$ presumably is an i.i.d. sequence. To this end, notice that the inference about unit roots for the models in H_{a1}^{a}, H_{a2}^{a}, and H_{a3}^{a} is now based on the NADF testing equations in $H_{a1}^{a,aux}$, $\tilde{H}_{a2}^{a,aux}$, and $\tilde{H}_{a3}^{a,aux}$, respectively.

2.3.3 The NPADF testing equations

The NADF testing equations presented above contain rather many regressors and increase rapidly, especially for $m = 1,2$, with p. It is therefore interesting to consider partially augmented testing equations. For $m = 1,2$, obvious

choices are to consider the following NPADF testing equations

$$H_{a1}^{pa,aux} \quad : \quad y_t = \mathbf{s}_{1t}'\boldsymbol{\lambda}_1^{pa} + (y_{t-1}\mathbf{s}_{1t})'\boldsymbol{\rho}_1^{pa} + \Delta\mathbf{y}_t'\boldsymbol{\zeta}_1^{pa} + u_t^*, \tag{2.20}$$

$$\tilde{H}_{a2}^{pa,aux} \quad : \quad y_t = \mathbf{s}_{2t}'\boldsymbol{\lambda}_2^{pa*} + (\xi_{t-1}\mathbf{s}_{1t})'\boldsymbol{\rho}_2^{pa} + \Delta\boldsymbol{\xi}_t'\boldsymbol{\zeta}_2^{pa} + u_t^*, \tag{2.21}$$

where $\boldsymbol{\lambda}_1^{pa} = (\lambda_{10}^{pa}, ..., \lambda_{13}^{pa})'$, $\boldsymbol{\lambda}_2^{pa*} = (\lambda_{20}^{pa*}, ..., \lambda_{24}^{pa*})'$, $\boldsymbol{\rho}_m^{pa} = (\rho_{m0}^{pa}, ..., \rho_{m3}^{pa})'$, $\boldsymbol{\zeta}_m^{pa} = (\zeta_{m01}^{pa}, \zeta_{m02}^{pa}, ..., \zeta_{m0,p-1}^{pa})'$, and u_t^* is an error term where $u_t^* = u_t$ holds when $\gamma = 0$. By (2.20) and (2.21) we keep the first two parts on the right-hand side of (2.12) and (2.16) (i.e. keeping the same set of regressors that are obtained in the case $p = 1$), but all terms with lagged differences of the dependent variables that are multiplied by a time trend are excluded. Using this approach, the number of reduced regressors equals $3(p-1)$. It is also evident that with this choice of NPADF testing equations, the equation in (2.17) qualifies both as a NADF and NPADF testing equation. Furthermore, the relevant null hypotheses are given by

$$\begin{aligned} H_{01}^{pa,aux} : \quad & \lambda_{1i}^{pa} = 0 \ \ \forall i, \ \rho_{10}^{pa} = 1, \ \rho_{1j}^{pa} = 0, \ j \geq 1, \\ & \zeta_{10q}^{pa} \in \mathbb{R}, \ q = 1, ..., p-1, \end{aligned} \tag{2.22}$$

$$\begin{aligned} \tilde{H}_{02}^{pa,aux} : \quad & \lambda_{21}^{pa*} \in \mathbb{R}, \ \lambda_{2i}^{pa*} = 0, \ i \neq 1, \ \rho_{20}^{pa} = 1, \ \rho_{2j}^{pa} = 0, \ j \geq 1, \\ & \zeta_{20q}^{pa} \in \mathbb{R}, \ q = 1, ..., p-1. \end{aligned} \tag{2.23}$$

Imposing the restrictions in (2.20) and (2.21) on (2.22) and (2.23), respectively, we obtain the same DGP's under the null hypotheses as in Section 2.3.2, i.e. $y_t = y_{t-1} + v_t$ and $\xi_t = \xi_{t-1} + v_t$, respectively.

2.4 The NDF, NADF, and NPADF unit root tests

In this section several unit root tests based on the NDF, NADF, and NPADF testing equations in the previous section are derived. The unit root tests will therefore be denoted as the NDF, NADF, and NPADF unit root tests.

All theoretical results presented below are derived under the assumption that the errors u_t in (2.1) are i.i.d. yielding limiting distributions for the test statistics that are nuisance parameter free (or that the nuisance parameters are trivially eliminated). Although, similar results can be established under the assumption that the error term in (2.1) is a strong mixing, the inference will be affected by nuisance parameters depending on the mixing assumption, see e.g. Phillips (1987), Phillips and Perron (1988), and Chapter 1. As a remark on notation in the following theorems and corollaries, $\xrightarrow{d}$ and $\xrightarrow{p}$ denote convergence in distribution and probability, respectively, and $B(r)$ denotes a standard Brownian motion defined on $[0, 1]$.

Assumption 1 *Let $\{u_t\}_{t=1}^{\infty}$ be an i.i.d. sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $E(u_t) = 0$, $E(u_t^2) = \sigma_u^2$, and $E(u_t^4) < \infty$ hold for all t.*

Theorem 2.1 *Consider models (2.6), (2.8), and (2.9) when (2.7), (2.10), and (2.11) hold, respectively. Furthermore, assume that $\{u_t\}_{t=1}^{\infty}$ fulfills Assumption 1. Then, for $m = 1, 2, 3$,*

$$\gamma_{m1}(\hat{\boldsymbol{\psi}}_m - \boldsymbol{\psi}_m) \xrightarrow{d} \boldsymbol{\Psi}_m^{-1}\boldsymbol{\Pi}_m, \tag{2.24}$$

and

$$\hat{\boldsymbol{\psi}}_m - \boldsymbol{\psi}_m \xrightarrow{p} \mathbf{0},$$

where

$$\begin{aligned}
\gamma_{11} &= diag\{\mathbf{T}_1\},\ \mathbf{T}_1 = \left[\ T^{1/2}\ \ T^{3/2}\ \ T^{5/2}\ \ T^{7/2}\ \ T\ \ T^2\ \ T^3\ \ T^4\ \right],\\
\gamma_{21} &= diag\{\mathbf{T}_2\},\ \mathbf{T}_2 = \left[\ T^{1/2}\ \ T^{3/2}\ \ T^{5/2}\ \ T^{7/2}\ \ T^{9/2}\ \ T\ \ T^2\ \ T^3\ \ T^4\ \right],\\
\gamma_{31} &= diag\{\mathbf{T}_3\},\ \mathbf{T}_3 = \left[\ T^{1/2}\ \ T^{3/2}\ \ T^{5/2}\ \ T^{7/2}\ \ T^{9/2}\ \ T\ \right],
\end{aligned}$$

and $\hat{\boldsymbol{\psi}}_1 = \left(\hat{\boldsymbol{\lambda}}_1', \hat{\boldsymbol{\varphi}}_1'\right)'$, $\hat{\boldsymbol{\psi}}_2 = \left(\hat{\boldsymbol{\lambda}}_2^{*\prime}, \hat{\boldsymbol{\varphi}}_2'\right)'$, and $\hat{\boldsymbol{\psi}}_3 = \left(\hat{\boldsymbol{\lambda}}_3^{*\prime}, \hat{\varphi}_{30}\right)'$ are the least square estimators of $\boldsymbol{\psi}_1 = (\boldsymbol{\lambda}_1', \boldsymbol{\varphi}_1')'$, $\boldsymbol{\psi}_2 = (\boldsymbol{\lambda}_2', \boldsymbol{\varphi}_2')'$, and $\boldsymbol{\psi}_3 = (\boldsymbol{\lambda}_3', \boldsymbol{\varphi}_3')'$ in (2.6), (2.8), and (2.9), respectively. Moreover, in (2.24),

$$\boldsymbol{\Psi}_m = \begin{bmatrix} \mathbf{A}_m & \sigma_u \mathbf{B}_m \\ \sigma_u \mathbf{B}_m' & \sigma_u^2 \mathbf{C}_m \end{bmatrix}, \quad \boldsymbol{\Pi}_m = \begin{bmatrix} \sigma_u \mathbf{D}_m \\ \sigma_u^2 \mathbf{E}_m \end{bmatrix},$$

where

$$\begin{array}{llllll}
\mathbf{A}_1 = [a_{ij}]_{4\times 4}, & \mathbf{B}_1 = [b_{ij}]_{4\times 4}, & \mathbf{C}_1 = [c_{ij}]_{4\times 4}, & \mathbf{D}_1 = [d_i]_{4\times 1}, & \mathbf{E}_1 = [e_i]_{4\times 1}, \\
\mathbf{A}_2 = [a_{ij}]_{5\times 5}, & \mathbf{B}_2 = [b_{ij}]_{5\times 4}, & \mathbf{C}_2 = \mathbf{C}_1, & \mathbf{D}_2 = [d_i]_{5\times 1}, & \mathbf{E}_2 = \mathbf{E}_1, \\
\mathbf{A}_3 = \mathbf{A}_2, & \mathbf{B}_3 = [b_{ij}]_{5\times 1}, & \mathbf{C}_3 = [c_{ij}]_{1\times 1}, & \mathbf{D}_3 = \mathbf{D}_2, & \mathbf{E}_3 = [e_i]_{1\times 1},
\end{array}$$

with

$$\begin{aligned}
a_{ij} &= 1/(i+j-1),\\
b_{ij} &= \int_0^1 r^{i+j-2} B(r)dr,\\
c_{ij} &= \int_0^1 r^{i+j-2} B(r)^2 dr,\\
d_i &= B(1) - (i-1)\int_0^1 r^{i-2} B(r)dr,\\
e_i &= \left(B(1)^2 - (i-1)\int_0^1 r^{i-2} B(r)^2 dr - 1/i\right)/2.
\end{aligned}$$

Proof. See Appendix A. ■

The joint limiting distributions in (2.24) contain the nuisance parameter σ_u. However, it is clear that the sub-matrices $\mathbf{A}_m$, $\mathbf{B}_m$, $\mathbf{C}_m$, $\mathbf{D}_m$, and $\mathbf{E}_m$ are nuisance parameter free.

Corollary 2.2 *(NDF tests) Suppose that the conditions of Theorem 2.1 hold. Define the test statistic*

$$\psi_m \equiv T(\hat{\varphi}_{m0} - 1), \quad m = 1, 2, 3.$$

Then,

$$\psi_m \xrightarrow{d} \mathbf{r}_m \tilde{\mathbf{\Psi}}_m^{-1} \tilde{\mathbf{\Pi}}_m, \tag{2.25}$$

where $\mathbf{r}_1 = [r_{1i}]_{1\times 8}$,$\mathbf{r}_2 = [r_{2i}]_{1\times 9}$,*and* $\mathbf{r}_3 = [r_{3i}]_{1\times 6}$, *with elements defined as* $r_{15} = r_{26} = r_{36} = 1$ *and zero otherwise. Moreover, in (2.25),*

$$\tilde{\mathbf{\Psi}}_m = \begin{bmatrix} \mathbf{A}_m & \mathbf{B}_m \\ \mathbf{B}_m' & \mathbf{C}_m \end{bmatrix}, \quad \tilde{\mathbf{\Pi}}_m = \begin{bmatrix} \mathbf{D}_m \\ \mathbf{E}_m \end{bmatrix}.$$

Proof. See Appendix A. ■

Corollary 2.3 *(NDF t-type of tests) Suppose that the conditions of Theorem 2.1 hold. Define the t-test statistic*

$$t_m \equiv (\hat{\varphi}_{m0} - 1)/\hat{\sigma}_{\hat{\varphi}_{m0}}, \quad m = 1, 2, 3,$$

where $\hat{\sigma}_{\hat{\varphi}_{m0}}$ *is the estimated standard deviations of* $\hat{\varphi}_{m0}$. *Then*

$$t_m \xrightarrow{d} \frac{\mathbf{r}_m \tilde{\mathbf{\Psi}}_m^{-1} \tilde{\mathbf{\Pi}}_m}{\left(\mathbf{r}_m \tilde{\mathbf{\Psi}}_m^{-1} \mathbf{r}_m'\right)^{1/2}}. \tag{2.26}$$

Proof. See Appendix A. ■

It is evident from (2.25) and (2.26) in Corollary 2.2 and 2.3, respectively, that both the NDF and the NDF t-type of tests have asymptotic distributions that are nuisance parameter free. Another important property is that for $m = 2, 3$, the tests in (2.25) and (2.26) are invariant with respect to if the random walk under the null hypothesis is with or without drift, i.e. the tests are invariant with respect to the parameters λ_{20} and λ_{30}, respectively. For $m = 1$ this is not the case. Moreover, it is only the test based on the LSTD (1) model that are invariant with respect to maintained trends up to order three under the null hypothesis, as noted in Ouliaris, Park, and Phillips (1989). In the present context this means that for $m = 3$, the NDF and the NDF t-type

of tests are invariant with respect to $\lambda_{30}, ..., \lambda_{33}$. However, in our case the restrictions $\lambda_{31} = ... = \lambda_{33} = 0$ in H_{03}^{aux} are implied restrictions from those that are initially imposed on the LSTD(1) model. Even though the restrictions in this case are not needed in the testing equation in H_{a3}^{aux}, it seems rather *ad hoc* to relax them in terms of the originally state null hypothesis in (2.4).

Assumption 2 *Let $\{u_t\}_{t=1}^{\infty}$ be an i.i.d. sequence of random variables satisfying Assumption 1. Furthermore, define $v_t \equiv c(L)u_t = \sum_{j=0}^{\infty} c_j u_{t-j}$, where $c(L) = \sum_{j=0}^{\infty} c_j L^j$ is a one-sided moving average polynomial in the lag operator such that (A.2.1) $c(1) \neq 0$ (no unit roots), and (A.2.2) $\sum_{m=1}^{\infty} j|c_j| < \infty$ (one-summability).*

Due to the earlier assumptions of a single unit root, (A.2.1) in Assumption 2 is satisfied under $H_{0m}^{a,aux}$ The only assumption we have to impose is (A.2.2), i.e. we have to control for the amount of serial dependence in $\{v_t\}_{t=1}^{\infty}$.

Theorem 2.4 *Consider models (2.12), (2.16), and (2.17) when (2.15), (2.18), and (2.19) hold. Furthermore, assume that $\{u_t\}_{t=1}^{\infty}$ and $\{v_t\}_{t=1}^{\infty}$ fulfill Assumption 2. Then, for $m = 1, 2, 3$,*

$$\gamma_{m2}^{a}(\hat{\psi}_m^a - \psi_m^a) \xrightarrow{d} (\mathbf{\Psi}_m^a)^{-1} \mathbf{\Pi}_m^a, \tag{2.27}$$

and

$$\hat{\psi}_m^a - \psi_m^a \xrightarrow{p} \mathbf{0},$$

where

$$\begin{aligned}
\gamma_{12}^a &= diag\left\{ \mathbf{T}_1, \ \left([\ T^{1/2} \ \ T^{3/2} \ \ T^{5/2} \ \ T^{7/2}\ ]' \otimes \mathbf{1} \right)' \right\}, \\
\gamma_{22}^a &= diag\left\{ \mathbf{T}_2, \ \left([\ T^{1/2} \ \ T^{3/2} \ \ T^{5/2} \ \ T^{7/2}\ ]' \otimes \mathbf{1} \right)' \right\}, \\
\gamma_{32}^a &= diag\left\{ \mathbf{T}_3, \ \left(T^{1/2}\mathbf{1} \right)' \right\},
\end{aligned}$$

and $\mathbf{1}$ is a $(p-1)\times 1$ vector of ones, $\hat{\psi}_1^a = \left(\hat{\lambda}_1^{a\prime}, \hat{\rho}_1^{a\prime}, \hat{\zeta}_1^{a\prime}\right)'$, $\hat{\psi}_2^a = \left(\hat{\lambda}_2^{a*\prime}, \hat{\rho}_2^{a\prime}, \hat{\zeta}_2^{a\prime}\right)'$, and $\hat{\psi}_3^a = \left(\hat{\lambda}_3^{a*\prime}, \hat{\rho}_{30}^{a\prime}, \hat{\zeta}_3^{a\prime}\right)'$ are the least square estimators of $\psi_1^a = \left(\lambda_1^{a\prime}, \rho_1^{a\prime}, \zeta_1^{a\prime}\right)'$, $\psi_2^a = \left(\lambda_2^{a*\prime}, \rho_2^{a\prime}, \zeta_2^{a\prime}\right)'$, and $\psi_3^a = \left(\lambda_3^{a*\prime}, \rho_{30}^{a\prime}, \zeta_3^{a\prime}\right)'$ in (2.12), (2.16), and (2.17), respectively. Moreover, in (2.27),

$$(\mathbf{\Psi}_m^a)^{-1} \mathbf{\Pi}_m^a = \begin{bmatrix} \mathbf{P}_m^a \\ \mathbf{Q}_m^a \end{bmatrix},$$

with sub-matricies given by

$$\mathbf{P}_m^a = \begin{bmatrix} \mathbf{A}_m & \lambda\mathbf{B}_m \\ \lambda\mathbf{B}_m' & \lambda^2\mathbf{C}_m \end{bmatrix}^{-1} \begin{bmatrix} \sigma_u\mathbf{D}_m \\ \sigma_u\lambda\mathbf{E}_m \end{bmatrix}, \quad \mathbf{Q}_m^a \sim N\left(\mathbf{0},\sigma_u^2\mathbf{V}_m^{-1}\right),$$

where $\lambda = c(1)\sigma_u$, *and*

$$\mathbf{V}_1 = \mathbf{A}_1 \otimes \mathbf{\Sigma}_1, \quad \mathbf{V}_2 = \mathbf{A}_1 \otimes \mathbf{\Sigma}_2, \quad \mathbf{V}_3 = \mathbf{\Sigma}_3,$$

with

$$\mathbf{\Sigma}_m = \begin{bmatrix} \gamma_{m,0} & \gamma_{m,1} & \cdots & \gamma_{m,p-2} \\ \gamma_{m,1} & \gamma_{m,0} & \cdots & \gamma_{m,p-3} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,p-2} & \gamma_{m,p-3} & \cdots & \gamma_{m,0} \end{bmatrix},$$

where $\gamma_{m,j}$ *equals* $E\left[\Delta y_t \Delta y_{t-j}\right]$ *for* $m = 1$ *and* $E\left[\Delta\xi_t\Delta\xi_{t-j}\right]$ *for* $m = 2, 3$.

Proof. See Appendix A. ■

Theorem 2.5 *Consider models (2.20) and (2.21) when (2.22) and (2.23) hold, respectively. Furthermore, assume that* $\{u_t\}_{t=1}^{\infty}$ *and* $\{v_t\}_{t=1}^{\infty}$ *fulfill Assumption 2. Then, for* $m = 1, 2, 3$,

$$\gamma_{m2}^{pa}(\hat{\psi}_m^{pa} - \psi_m^{pa}) \xrightarrow{d} (\mathbf{\Psi}_m^{pa})^{-1}\,\mathbf{\Pi}_m^{pa}, \tag{2.28}$$

and

$$\hat{\psi}_m^{pa} - \hat{\psi}_m^{pa} \xrightarrow{p} \mathbf{0},$$

where

$$\gamma_{m2}^{pa} = diag\left\{ \mathbf{T}_m \quad (T^{1/2}\mathbf{1})' \right\},$$

and $\hat{\psi}_1^{pa} = \left(\hat{\boldsymbol{\lambda}}_1^{pa\prime}, \hat{\boldsymbol{\rho}}_1^{pa\prime}, \hat{\boldsymbol{\zeta}}_1^{pa\prime}\right)'$ *and* $\hat{\psi}_2^{pa} = \left(\hat{\boldsymbol{\lambda}}_2^{pa*\prime}, \hat{\boldsymbol{\rho}}_2^{pa\prime}, \hat{\boldsymbol{\zeta}}_2^{pa\prime}\right)'$ *are the least square estimators of* $\psi_1^{pa} = \left(\boldsymbol{\lambda}_1^{pa\prime}, \boldsymbol{\rho}_1^{pa\prime}, \boldsymbol{\zeta}_1^{pa\prime}\right)'$ *and* $\hat{\psi}_2^{pa} = \left(\hat{\boldsymbol{\lambda}}_2^{pa*\prime}, \hat{\boldsymbol{\rho}}_2^{pa\prime}, \hat{\boldsymbol{\zeta}}_2^{pa\prime}\right)'$ *in (2.20) and (2.21), respectively. Moreover, in (2.28),*

$$(\mathbf{\Psi}_m^{pa})^{-1}\,\mathbf{\Pi}_m^{pa} = \begin{bmatrix} \mathbf{P}_m^{pa} \\ \mathbf{Q}_m^{pa} \end{bmatrix},$$

where

$$\mathbf{P}_m^{pa} = \mathbf{P}_m^a, \quad \mathbf{Q}_m^{pa} \sim N\left(\mathbf{0},\sigma_u^2\mathbf{\Sigma}_m^{-1}\right).$$

Proof. See Appendix A. ■

Note first that we allow for $m = 3$ in Theorem 2.5. This is only for notational convenience due to the classification about regression models made in Section 2.3.3, and thus it follows that $\boldsymbol{\gamma}_{32}^{pa}(\hat{\boldsymbol{\psi}}_3^{pa} - \boldsymbol{\psi}_3^{pa}) = \boldsymbol{\gamma}_{32}^{a}(\hat{\boldsymbol{\psi}}_3^{a} - \boldsymbol{\psi}_3^{a})$, and $(\boldsymbol{\Psi}_3^a)^{-1}\boldsymbol{\Pi}_3^a$ in Theorem 2.4 equals $(\boldsymbol{\Psi}_3^{pa})^{-1}\boldsymbol{\Pi}_3^{pa}$ in Theorem 2.5. Furthermore, both the asymptotic distributions (2.27) and (2.28) in Theorems 2.4 and 2.5 contain the nuisance parameters σ_u and λ.

Corollary 2.6 *(NADF and NPADF type of tests) Suppose that the conditions of Theorems 2.4 and 2.5 hold. Define the test statistics,*

$$\psi_m^a \equiv T(\hat{\rho}_{m0}^a - 1), \quad \psi_m^{pa} \equiv T(\hat{\rho}_{m0}^{pa} - 1), \quad m = 1, 2, 3.$$

Then,

$$c^* \psi_m^a \xrightarrow{d} \mathbf{r}_m \tilde{\boldsymbol{\Psi}}_m^{-1} \tilde{\boldsymbol{\Pi}}, \tag{2.29}$$

and

$$c^* \psi_m^{pa} \xrightarrow{d} \mathbf{r}_m \tilde{\boldsymbol{\Psi}}_m^{-1} \tilde{\boldsymbol{\Pi}}, \tag{2.30}$$

where $c^* = \lambda/\sigma_u$.

Proof. See Appendix A. ■

Especially three interesting properties about the NADF and NPADF type of tests are revealed in Corollary 2.6. First, although the tests ψ_m^a and ψ_m^{pa} allow for serially correlated errors, their limiting distributions are trivially adjusted to become nuisance parameter free. As such, we instead obtain a test statistic depending on nuisance parameters through the quantity c^*. Second, the tests $c^*\psi_m^a$ based on the NADF testing equations have the same asymptotic distribution as the $c^*\psi_m^{pa}$ tests based on the NPADF testing equations. Even though $c^*\psi_m^a$ is a test involving the estimation of $3(p-1)$ more regressors than the $c^*\psi_m^{pa}$ test, the extra set of regressors do not affect the asymptotic distributions. Third, the asymptotic distributions for $c^*\psi_m^a$ and $c^*\psi_m^{pa}$ collapse into the same limiting distributions as for the ψ_m tests with $p = 1$ in Corollary 2.2. This means that the tests $c^*\psi_m^a$ and $c^*\psi_m^{pa}$ are asymptotically invariant with respect to the number of lags included in the NADF and NPADF testing equations.

Corollary 2.7 *(NADF and NPADF t-type of tests) Suppose that the conditions of Theorem 2.4 and 2.5 hold. Define the t-test statistics,*

$$t_m^a \equiv (\hat{\rho}_{m0}^a - 1)/\hat{\sigma}_{\hat{\rho}_{m0}^a}, \quad t_m^{pa} \equiv (\hat{\rho}_{m0}^{pa} - 1)/\hat{\sigma}_{\hat{\rho}_{m0}^{pa}}, \quad m = 1, 2, 3,$$

where $\hat{\sigma}_{\hat{\rho}^a_{m0}}$ *and* $\hat{\sigma}_{\hat{\rho}^{pa}_{m0}}$ *are the estimated standard deviations of* $\hat{\rho}^a_{m0}$ *and* $\hat{\rho}^{pa}_{m0}$*, respectively. Then,*

$$t^a_m \xrightarrow{d} \frac{\mathbf{r}_m \tilde{\mathbf{\Psi}}_m^{-1} \tilde{\mathbf{\Pi}}_m}{\left(\mathbf{r}_m \tilde{\mathbf{\Psi}}_m^{-1} \mathbf{r}'_m\right)^{1/2}}, \tag{2.31}$$

and

$$t^{pa}_m \xrightarrow{d} \frac{\mathbf{r}_m \tilde{\mathbf{\Psi}}_m^{-1} \tilde{\mathbf{\Pi}}_m}{\left(\mathbf{r}_m \tilde{\mathbf{\Psi}}_m^{-1} \mathbf{r}'_m\right)^{1/2}}.$$

Proof. See Appendix A. ■

Corollary 2.7 reveals that the NADF and NPADF t-type of tests mainly have the same asymptotic properties as the NADF and NPADF type of tests in Corollary 2.6. However, in contrast to the tests in Corollary 2.6, the limiting distributions for the tests statistics t^a_m and t^{pa}_m need no further manipulations to become nuisances parameter free.

2.5 Asymptotic and finite-sample properties of the nonlinear Dickey-Fuller type of tests

In this section we present the asymptotic as well as finite-sample critical values for some of the tests presented in Section 2.4. Due to the findings in Corollaries 2.6 and 2.7, the asymptotic critical values for the tests considering partially or fully augmented models are the same. However, from a finite-sample point of view, we choose only to present the critical values for the tests based on partially augmented models since they require estimation of less parameters. Moreover, a size study is performed for a NPADF t-type of test. Finally, the power properties of the tests based on the LSTAR and the LSTART models are examined.

2.5.1 Asymptotic and finite-sample critical values

When generating the asymptotic distribution of the tests ψ_m, t_m, $c^*\psi^{pa}_m$, and t^{pa}_m, we let $T = 1\ 000\ 000$ to simulate a Brownian motion $B(r)$ on $[0, 1]$, and the number of replications are set to 1 000 000. The finite-sample critical values for the tests ψ_m and t_m are obtained by simulating data from the model $y_t = y_{t-1} + u_t$ where $u_t \sim nid(0, 1)$ with desired sample sizes, and thereafter the test statistics in Corollaries 2.2 and 2.3 are calculated. This procedure is repeated 1 000 000 times, yielding the finite-sample distributions of the tests.

Furthermore, when generating the finite-sample critical values for the test statistics $c^*\psi_m^{pa}$ and t_m^{pa}, we proceed as above but we simulate data from the model $\left(1-\zeta_{m01}^{pa}-\ldots-\zeta_{m,0,p-1}^{pa}\right)\Delta y_t = u_t$. Asymptotically this causes no problems, which is shown in Corollaries 2.6 and 2.7, but all finite-sample distributions depend on the nuisance parameters $\zeta_{m01}^{pa},\ldots,\zeta_{m0,p-1}^{pa}$. To proceed we let $\zeta_{m0q}^{pa}=0$ for all m and q under the null hypothesis, and in the next subsection it is shown that the tests and their finite-sample distributions are rather robust against the values of ζ_{m0q}^{pa}, and the size distortions of the tests become modest. To operationalize $c^*\psi_m^{pa}$ we note that $c^* = 1/\left(1-\zeta_{m01}^{pa}-\ldots-\zeta_{m,0,p-1}^{pa}\right)$, which is consistently estimated by $1/\left(1-\hat{\zeta}_{m01}^{pa}-\ldots-\hat{\zeta}_{m,0,p-1}^{pa}\right)$.[4]

From Table 2.1 it is seen that the critical values for ψ_m are large in absolute values and negative. The same conclusions are drawn for t_m, however, with critical values being standardized. In fact, the empirical distributions turn out to be heavily skewed to the left with negative means. The reason for this is the inclusion of time trends of high orders creating a downward bias in $\hat{\varphi}_{m0}$. In addition, for $m=1,2$ and the ψ_m tests, there are big differences for the critical values among different sample sizes. Furthermore, the critical values for the ψ_2 test are the same as for the T_3 test in Chapter 1 and they are there presented in the Table 1.1. In addition, the critical values for the ψ_3 test and the sample size $T=500$, are the same critical values as for the $K_4(\hat{\alpha})$ test in Ouliaris, Park, and Phillips (1989) and are reported in their Appendix 2.

In Table 2.2 the critical values for the tests $c^*\psi_m^{pa}$ and t_m^{pa} with $p=2$ are presented. Critical values for $p=3,4$ can be found in Tables B.1 and B.2 in the Appendix B. Notable is that in small samples the estimates of $\hat{\zeta}_{m01}$ on average are quite inaccurate. This, of course, affects the finite-sample distributions because c^* is biased upwards and thus the critical values become even more negative than corresponding critical values in Table 2.1. However, as the sample size tends to infinity, the critical values at all significance levels for $c^*\psi_m^{pa}$ and t_m^{pa} equal the critical values for ψ_m and t_m in Table 2.1 and $T=\infty$, which also confirms the findings in Corollaries 2.6 and 2.7.

As a final remark, the critical values presented in Tables 2.1 and 2.2 can be seen as the critical values for unit root tests in LSTAR(p), LSTART(p), and LSTD(p) models with a transition function allowing for three transitions, i.e.

$$F(t) = \frac{1}{1+\exp\{-\gamma(t-c_1)(t-c_2)(t-c_3)\}} - \frac{1}{2},$$

and where we instead apply a first-order Taylor approximation. The reason

[4]In our set-up this means that as $T\to\infty$, $\hat{\zeta}_{m01}^{pa}\to 0,\ldots,\hat{\zeta}_{m,0,p-1}^{pa}\to 0$ hold for all m.

Table 2.1 Critical values for the NDF tests ψ_m and t_m in Corollaries 2.2 and 2.3.

	ψ_1			t_1		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-109.61	-81.81	-68.15	-3.54	-2.82	-2.41
100	-125.97	-92.91	-76.83	-3.64	-2.93	-2.54
250	-142.22	-102.54	-83.87	-3.75	-3.00	-2.61
500	-149.16	-106.44	-86.57	-3.76	-3.04	-2.65
1000	-152.84	-108.44	-87.93	-3.76	-3.04	-2.65
∞	-156.45	-109.40	-89.33	-3.77	-3.05	-2.66
	ψ_2			t_2		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-119.64	-90.06	-75.56	-3.40	-2.65	-2.26
100	-138.48	-103.72	-86.31	-3.53	-2.80	-2.41
250	-159.44	-116.44	-95.79	-3.58	-2.89	-2.52
500	-168.48	-121.80	-99.73	-3.63	-2.91	-2.54
1000	-173.38	-124.70	-101.80	-3.65	-2.94	-2.56
∞	-175.31	-126.50	-103.11	-3.66	-2.95	-2.57
	ψ_3			t_3		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-38.94	-32.69	-29.50	-5.38	-4.70	-4.36
100	-43.77	-35.90	-32.07	-5.20	-4.61	-4.30
250	-47.11	-38.13	-33.80	-5.11	-4.55	-4.26
500	-48.27	-38.90	-34.42	-5.08	-4.53	-4.26
1000	-48.31	-39.33	-34.77	-5.07	-4.52	-4.25
∞	-48.50	-39.83	-34.91	-5.05	-4.51	-4.24

Notes: The probability shown at the head of each column is the area in the left-hand tail. The results are based on 1 000 000 replications.

Table 2.2 Critical values for the NPADF tests $c^*\psi_m^{pa}$ and t_m^{pa} in Corollaries 2.6 and 2.7 with $p = 2$.

		$c^*\psi_1^{pa}$			t_1^{pa}		
T	$\hat{\zeta}_{101}$	0.01	0.05	0.10	0.01	0.05	0.10
50	0.160	-184.30	-127.99	-103.68	-3.90	-3.11	-2.70
100	0.081	-161.97	-115.78	-94.54	-3.82	-3.08	-2.68
250	0.033	-157.32	-112.16	-91.29	-3.80	-3.07	-2.67
500	0.017	-156.66	-111.33	-90.41	-3.78	-3.06	-2.67
1000	0.008	-156.50	-110.40	-90.01	-3.77	-3.05	-2.66
∞	0.000	-156.46	-109.5	-89.89	-3.77	-3.05	-2.66
		$c^*\psi_2^{pa}$			t_2^{pa}		
T	$\hat{\zeta}_{201}$	0.01	0.05	0.10	0.01	0.05	0.10
50	0.180	-211.89	-148.73	-121.31	-3.80	-3.01	-2.60
100	0.095	-185.66	-133.45	-109.68	-3.72	-2.99	-2.59
250	0.039	-178.53	-128.80	-105.62	-3.69	-2.97	-2.58
500	0.020	-178.00	-128.11	-104.74	-3.68	-2.97	-2.58
1000	0.010	-177.90	-127.63	-103.90	-3.67	-2.96	-2.57
∞	0.000	-175.60	-126.55	-103.15	-3.66	-2.95	-2.57
		$c^*\psi_3^{pa}$			t_3^{pa}		
T	$\hat{\zeta}_{301}$	0.01	0.05	0.10	0.01	0.05	0.10
50	0.127	-66.05	-50.19	-43.21	-5.42	-4.72	-4.38
100	0.064	-56.91	-44.37	-38.71	-5.22	-4.61	-4.31
250	0.025	-52.30	-41.50	-36.40	-5.11	-4.55	-4.26
500	0.013	-50.88	-40.67	-35.81	-5.08	-4.54	-4.25
1000	0.006	-50.19	-40.16	-35.41	-5.07	-4.53	4.25
∞	0.000	-48.70	-39.87	-34.99	-5.06	-4.51	-4.24

Notes: The probability shown at the head of each column is the area in the left-hand tail. The reported estimates of the nuisance parameters are average OLS estimates from 1 000 000 replications when the true values are set to zero.

for this is that in the present case the same set of explanatory variables as in the cases H_{a1}^{aux}-H_{a3}^{aux}, $H_{a1}^{a,aux}$-$H_{a3}^{a,aux}$, and $H_{a1}^{pa,aux}$-$H_{a3}^{pa,aux}$, are obtained, and the same null hypotheses will be tested.

2.5.2 A size study

We examine the size properties of the t_1^{pa} test statistic with $p = 2$ and the influence of the nuisance parameter ζ_{101}. The results are reported in Table 2.3. The size properties of the tests t_2^{pa} and t_3^{pa} are similar and therefore excluded. Moreover, in small samples the t_m^{pa} tests are preferable to the ψ_m^{pa} type of tests, and the main reason for this is, though not reported here, that in general the size distortions are smaller for the t_m^{pa} tests.[5]

In Table 2.3 we see that the size distortions are substantial, and this especially for $|\zeta_{101}| \geq 0.8$. The size distortions are mitigated by increasing the sample size, and become negligible for $T \geq 500$ except in the two extreme cases $\zeta_{101} = \pm 0.99$. When the estimated size is larger than 6%, we choose to report size-adjusted critical values, given within the parentheses in Table 2.3.

Another source of size distortion could arise if we relax our assumption on the error term v_t and allow for some other error structure. It is well known that the classical Dickey and Fuller type of tests suffer from large size distortion (up to 90%) when the error term has a unit root moving average structure, see e.g. Schwert (1989) and Chapter 1. A natural approach to this problem is to allow for a wider class of error terms, e.g. by assuming that the error term is a strong mixing. However, as pointed out in Schwert (1989) and Chapter 1, this approach does not solve the problem satisfactorily, and the size distortions are not eliminated. It seems better to adopt the instrumental variable approach suggested by Hall (1989), or adding lags of the error term. The latter is an approach that is at hand. In fact, at a 5% nominal level the estimated size becomes 12% for the t_1^{pa} test when $T = 250$ and the error term is an MA(1) process with moving average coefficient equal to -0.8 (not reported here). This should be compared to 99% in a similar testing situation in Chapter 1. In general, increasing the order of augmentation further mitigates the size distortions.

2.5.3 Power studies

The power properties of the t_1 and t_2 tests in Corollary 2.3 are investigated. The findings for the t_3 test are similar and therefore omitted. It should be

[5]The ranking of the tests assumes that one would like to control for the significance levels of the tests. If power would be the criteria, the ψ_m^{pa} tests should, in general, be used.

Table 2.3 Empirical size of the NPADF t_1^{pa} test in Corollary 2.7 with $p = 2$.

	$T = 50$		$T = 100$		$T = 250$	
ζ_{101}	$\hat{\zeta}_{101}$	Size	$\hat{\zeta}_{101}$	Size	$\hat{\zeta}_{101}$	Size
0.99	0.59	0.115 (-3.64)	0.75	0.111 (-3.66)	0.88	0.109 (-3.61)
0.8	0.57	0.090 (-3.49)	0.68	0.079 (-3.36)	0.76	0.061 (-3.19)
0.6	0.50	0.066 (-3.30)	0.55	0.063 (-3.23)	0.58	0.053
0.4	0.40	0.056	0.40	0.055	0.40	0.054
0.2	0.28	0.052	0.24	0.052	0.22	0.052
0.0	0.16	0.051	0.08	0.049	0.03	0.050
-0.2	-0.03	0.048	-0.08	0.049	-0.15	0.050
-0.4	-0.10	0.050	-0.25	0.046	-0.33	0.047
-0.6	-0.24	0.048	-0.41	0.050	-0.52	0.049
-0.8	-0.39	0.052	-0.58	0.053	-0.70	0.051
-0.99	-0.57	0.086 (-3.42)	-0.76	0.082 (-3.39)	-0.89	0.078 (-3.36)

	$T = 500$		$T = 1000$	
ζ_{101}	$\hat{\zeta}_{101}$	Size	$\hat{\zeta}_{101}$	Size
0.99	0.94	0.103 (-3.53)	0.97	0.09 (-3.47)
0.8	0.78	0.058	0.79	0.055
0.6	0.59	0.051	0.60	0.047
0.4	0.40	0.048	0.40	0.051
0.2	0.21	0.054	0.20	0.052
0.0	0.02	0.049	0.01	0.051
-0.2	-0.18	0.049	-0.19	0.049
-0.4	-0.36	0.049	-0.38	0.050
-0.6	-0.56	0.050	-0.58	0.053
-0.8	-0.75	0.048	-0.78	0.048
-0.99	-0.94	0.076 (-3.33)	-0.96	0.070 (-3.22)

Notes: The nominal size of the test is 5%. The reported estimates of the nuisance parameters are the average OLS estimates from 1 000 000 replications where the true values are shown in the column to the left. Values in parentheses correspond to size-adjusted critical values.

pointed out that the power of the ψ_m tests are generally higher than for the t_m tests. However, as argued in the previous subsection, the ψ_m tests may suffer from large size distortions, and consequently, the power results can be misleading. Furthermore, we choose not to report the power results for the the augmented tests, but one can expect the main findings to coincide with the results for the t_m tests.

2.5.3.1 Empirical power for the NDF t_1 type of test when the DGP is an LSTAR(1) model

The power of the t_1 test is examined when the LSTAR(1) model in H_{a1} is the DGP. Without loss of generality the transition function is replaced, as a matter of convenience, by $\tilde{F} \equiv F + 1/2$. It is clear that $\tilde{F}(t) : \mathbb{R}_+ \to [0,1]$. In all coming experiments, we impose the Lagrange stability condition,

$$\pi_{11} \in (0,1)\,, \quad \pi_{11} + \pi_{21} \in (0,1)\,, \tag{2.32}$$

on the skeleton of the LSTAR(1) model to rule out unrealistic trajectories (unstable or explosive trajectories), see Tong (1990). The parameters of the LSTAR(1) model are assigned to the following values

$$\begin{aligned} \pi_{10} &= 0, \quad \pi_{20} \in \{1,2\}, \quad \gamma \in \{0.01, 0.1, 1\}, \\ c &\in \{0.5T,\ 0.25T,\ 0.75T\}, \quad T = 250, \end{aligned} \tag{2.33}$$

and the autoregressive coefficients in the linear and nonlinear part are assigned values according to the scenarios

$$\begin{aligned} &\text{Low-High:} && \pi_{11} = 0.10, && \pi_{21} \in \{0.75, 0.80, 0.85\}, \\ &\text{Medium-High:} && \pi_{11} = 0.40, && \pi_{21} \in \{0.45, 0.45, 0.55\}, \\ &\text{High-High:} && \pi_{11} = 0.75, && \pi_{21} \in \{0.10, 0.15, 0.20\}. \end{aligned} \tag{2.34}$$

By this design we note that all trajectories start at zero and head towards the long-run attractor $\pi_{20}/(1-\pi_{11}-\pi_{21})$, as time evolves. Now varying π_{20}, as well as π_{11} and π_{21}, we see that the mean of the LSTAR(1) model at the end of the period ranges from 6.67 to 40 (assuming that a complete transition takes place). How fast the trajectories reach the long-run attractor is determined through γ. When $\gamma = 0.01$ the transition is slow and we encounter an almost

linear LSTAR(1) model.[6] Letting $\gamma = 0.1$, the transition is intermediate, although a sample size of 250 is enough for a complete transition to take place. When $\gamma = 1.00$ the transition is more abrupt and the LSTAR(1) model behaves almost like a TAR(1) process. We also vary the point in time around which the transition will take place. For instance, when $c = 0.25T$, it is illustrative to think that we have an TAR(1) model where the first 25% of the sample is an AR(1) process characterized by π_{11}, and in the remaining 75% of the sample the AR(1) process is characterized by the parameter vector $(\pi_{20}, \pi_{11} + \pi_{21})'$.

Finally, the three scenarios reflect the impact of the nonlinear part in the LSTAR model in terms of dynamics. The Low-High scenario describes an LSTAR model with a low value of the autoregressive root in the beginning of the period and reaches a closer to unstable state, $\pi_{11} + \pi_{21} = 0.95$, at the end of the period. The nonlinear impact of dynamics is viewed as high. The Medium-High scenario presents an LSTAR model where the autoregressive root is intermediate in the beginning of the period and reaches a closer to unstable state at the end of the period. The magnitude of nonlinear dynamics is seen as medium. In the High-High scenario the LSTAR model displays an autoregressive root with a high value in the beginning of the period and reaches a state closer to unstableness at the end of the period. The nonlinear change in dynamics is perceived as low.

We choose to compare the power of the t_1 test with the power of the Dickey-Fuller (DF) t-type of test based on the model $y_t = a_0 + a_1 y_{t-1} + u_t$. This will give the opportunity to investigate the bias towards nonrejection for the DF tests when the DGP has a smooth change in levels and dynamics. The results are presented in Tables 2.4-2.6.

For $\gamma = 0.01$ it is seen in Tables 2.4-2.6 that the DF test is superior to the t_1 test. This is to be expected since the LSTAR(1) model appears almost linear. The modest power for the t_1 test is explained by the fact that it is penalized due to an ambiguous set of explanatory variables.

When $\gamma = 0.10$ we notify the collapse in power of the DF test under all scenarios. The reason for this is that a full transition actually takes place for all values of c, and that the speed of transition is fast enough to generate truly nonlinear models with a clear shift in levels. The power of the t_1 test is increased and is close to unity under the Low-High and Medium-High scenarios with $c = 0.25T$, $0.5T$, and all values of π_{20} and π_{21}. This illumi-

[6]When $\gamma = 0.01$ a full transition does not take place during the sample period. Notice that $\tilde{F}(t = T = 250; \gamma = 0.01, c = 62.5) = 0.87$, $\tilde{F}(t = T = 250; \gamma = 0.01, c = 125) = 0.78$, and $\tilde{F}(t = T = 250; \gamma = 0.01, c = 187.5) = 0.66$.

Table 2.4 (Low-High) Empirical power of the t_1 test and the corresponding Dickey-Fuller t-type of test.

$T = 250$		$c = 0.5T$				$c = 0.25T$				$c = 0.75T$			
		t_1		DF		t_1		DF		t_1		DF	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
	0.75	0.54	0.64	1.00	1.00	0.47	0.63	1.00	1.00	0.58	0.64	1.00	1.00
0.01	0.80	0.53	0.64	1.00	1.00	0.47	0.63	1.00	0.97	0.57	0.63	1.00	1.00
	0.85	0.53	0.64	1.00	1.00	0.48	0.64	1.00	0.47	0.57	0.64	1.00	1.00
	0.75	0.97	1.00	0.45	0.00	1.00	1.00	0.47	0.00	0.31	0.24	0.98	0.00
0.10	0.80	0.99	1.00	0.00	0.00	1.00	0.98	0.00	0.00	0.30	0.25	0.47	0.00
	0.85	1.00	1.00	0.00	0.00	0.99	0.95	0.00	0.00	0.30	0.28	0.01	0.00
	0.75	1.00	1.00	0.23	0.00	0.96	0.54	0.00	0.00	0.19	0.33	0.81	0.00
1.00	0.80	1.00	1.00	0.00	0.00	0.87	0.29	0.00	0.00	0.21	0.51	0.05	0.00
	0.85	1.00	1.00	0.00	0.00	0.63	0.14	0.00	0.00	0.34	0.86	0.00	0.00

Note: The nominal size of the tests is 5%, and the results are based on 10 000 replications.

Table 2.5 (Medium-High) Empirical power of the t_1 test and the corresponding Dickey-Fuller t-type of test.

$T = 250$		$c = 0.5T$				$c = 0.25T$				$c = 0.75T$			
		t_1		DF		t_1		DF		t_1		DF	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
	0.45	0.27	0.36	1.00	1.00	0.24	0.39	1.00	0.99	0.27	0.34	1.00	1.00
0.01	0.50	0.26	0.37	1.00	0.99	0.25	0.39	1.00	0.60	0.28	0.34	1.00	1.00
	0.55	0.26	0.38	1.00	0.80	0.27	0.40	0.94	0.01	0.27	0.35	1.00	1.00
	0.45	0.77	0.98	0.00	0.00	0.95	0.93	0.00	0.00	0.11	0.08	0.85	0.00
0.10	0.50	0.87	0.99	0.00	0.00	0.95	0.87	0.00	0.00	0.11	0.09	0.16	0.00
	0.55	0.94	0.99	0.00	0.00	0.90	0.78	0.00	0.00	0.11	0.10	0.01	0.00
	0.45	0.95	1.00	0.23	0.00	0.76	0.25	0.00	0.00	0.07	0.12	0.51	0.00
1.00	0.50	0.99	1.00	0.00	0.00	0.61	0.11	0.00	0.00	0.08	0.22	0.01	0.00
	0.55	1.00	1.00	0.00	0.00	0.40	0.06	0.00	0.00	0.13	0.55	0.00	0.00

Note: The nominal size of the tests is 5%, and the results are based on 10 000 replications.

Table 2.6 (High-High) Empirical power of the t_1 test and the corresponding Dickey-Fuller t-type of test.

$T = 250$		$c = 0.5T$				$c = 0.25T$				$c = 0.75T$			
		t_1		DF		t_1		DF		t_1		DF	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
	0.10	0.07	0.13	1.00	0.83	0.08	0.17	1.00	0.86	0.06	0.09	1.00	0.96
0.01	0.15	0.07	0.13	0.96	0.19	0.09	0.17	0.90	0.12	0.06	0.10	1.00	0.69
	0.20	0.09	0.15	0.76	0.00	0.10	0.17	0.20	0.00	0.07	0.11	0.96	0.18
	0.10	0.22	0.53	0.05	0.00	0.35	0.37	0.35	0.00	0.03	0.02	0.40	0.00
0.10	0.15	0.32	0.62	0.00	0.00	0.41	0.35	0.00	0.00	0.02	0.02	0.01	0.00
	0.20	0.46	0.68	0.00	0.00	0.44	0.35	0.00	0.00	0.02	0.02	0.00	0.00
	0.10	0.37	0.67	0.01	0.00	0.21	0.05	0.00	0.00	0.01	0.02	0.17	0.00
1.00	0.15	0.52	0.73	0.00	0.00	0.18	0.03	0.00	0.00	0.01	0.04	0.00	0.00
	0.20	0.65	0.77	0.00	0.00	0.15	0.02	0.00	0.00	0.01	0.10	0.00	0.00

Note: The nominal size of the tests is 5%, and the results are based on 10 000 replications.

nates how the power of the t_1 test responds to the timing of the transition and the magnitude of nonlinearities. It also confirms that the linearity part in the LSTAR(1) model dominates both when $c = 0.75T$ or the High-High scenario is concerned, and consequently the t_1 test performs less satisfactorily. Even though in the present context the nonlinear part and its dynamics are suppressed, the generated trajectories appear nonlinear enough and the power for the DF test is modest. It is also interesting to note that the t_1 test is very robust against changes in the intercept and levels, in contrast to tests based on first-order Taylor approximation, see Chapter 1 for the discussion about the level leverage effect.

For $\gamma = 1.00$ the main findings are the same as for $\gamma = 0.10$. Especially note that for $c = 0.25T$ and $\pi_{20} = 2$, the power of the test t_1 decreases to rather low levels due to the complex trajectories of the LSTAR model.

To this end, we conclude that if we are right about the timing of the transition in the sense that it takes place around the middle of the sample ($c = 0.5T$), the t_1 test has very satisfactory power and *increases* in π_{20} and π_{21}, as well as with γ, in all scenarios. An opposite relationship for the power holds for the DF test, and the power is in general close to zero or equal to zero.

Table 2.7 (Low-High) Empirical power of the t_2 test and the corresponding Phillips-Perron t-type of test.

$T = 250$		$\pi_{23} = 0.005$				$\pi_{23} = 0.0075$				$\pi_{23} = 0.01$			
		t_2		PP		t_2		PP		t_2		PP	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
	0.75	0.59	0.62	1.00	1.00	0.61	0.63	1.00	1.00	0.63	0.65	1.00	1.00
0.01	0.80	0.59	0.61	1.00	1.00	0.61	0.65	1.00	1.00	0.64	0.65	1.00	1.00
	0.85	0.60	0.63	1.00	1.00	0.64	0.67	1.00	1.00	0.66	0.69	1.00	1.00
	0.75	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
0.10	0.80	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.85	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.75	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
1.00	0.80	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.85	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00

Note: The nominal size of the tests is 5%, and the results are based on 10 000 replications.

2.5.3.2 Empirical power for the NDF t_2 type of test when the DGP is an LSTART(1) model

The power of the t_2 test presented in Corollary 2.3 is explored when the DGP is the LSTART(1) model in H_{a2}. The analysis is conducted under the same prerequisites as in Subsection 2.5.3, but the analysis of varying the timing of the transition is excluded and only the case $c = 0.5T$ is in focus. In addition to these prerequisites, the coefficients of the time trend are assigned the following values

$$\pi_{12} = 0.005, \quad \pi_{22} \in \{0.005, 0.0075, 0.01\}.$$

The (linearly time-dependent) long-run attractors implied by this set-up are given by $\pi_{20}/(1-\pi_{11}-\pi_{21}) + (\pi_{12}+\pi_{22})/(1-\pi_{11}-\pi_{21})t$, and the slope coefficient ranges from 0.0667 to 0.3 at the end of the sample period (assuming that a complete transition takes place). The power of the t_2 test is compared to the classical t-type of test in Phillips and Perron (1988) denoted PP, and is based on the model $y_t = a_0 + a_1 y_{t-1} + a_2 t + u_t$. This means that we will investigate the bias towards nonrejection for a linear trend-stationary model when the DGP accommodates a smooth shift in levels, dynamics, and trends. The power results are presented in Tables 2.7-2.9.

Table 2.8 (Medium-High) Empirical power of the t_2 test and the corresponding Phillips-Perron t-type of test.

$T = 250$		$\pi_{23} = 0.005$				$\pi_{23} = 0.0075$				$\pi_{23} = 0.01$			
		t_2		PP		t_2		PP		t_2		PP	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
	0.45	0.30	0.33	1.00	1.00	0.32	0.33	1.00	1.00	0.34	0.36	1.00	1.00
0.01	0.50	0.32	0.34	1.00	1.00	0.33	0.37	1.00	0.99	0.37	0.39	1.00	0.97
	0.55	0.33	0.37	0.99	0.92	0.37	0.40	0.93	0.71	0.39	0.42	0.73	0.38
	0.45	1.00	1.00	0.01	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
0.10	0.50	1.00	1.00	0.01	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.55	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.45	1.00	1.00	0.01	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
1.00	0.50	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.55	1.00	0.98	0.00	0.00	1.00	0.96	0.00	0.00	0.99	0.01	0.00	0.00

Note: The nominal size of the tests is 5%, and the results are based on 10 000 replications.

For $\gamma = 0.01$ it appears in Tables 2.7-2.9 that the PP test outperforms, in general, the t_2 test. This is because the LSTART(1) model acts like a linear trend stationary process. More interesting is the extraordinary robustness in the power of the t_2 test under the Low-High and Medium-High scenarios whenever $\gamma \geq 0.1$ and regardless of how the values of the other parameters are varied. That is, when the nonlinear impact is medium or high, it appears that the design of the t_2 test is very satisfactory, and the probability of rejecting the null hypothesis when it is wrong is unity. The opposite is true for the PP test and the lack of power is evident. Moreover, in the High-High scenario in Table 2.9, we see that the power is substantial but, however, lower than for the two other scenarios. This emphasizes that the power for the t_2 test is reduced due to a less evident impact of nonlinearity in the High-High scenario.

To this end, we note that the collapse in power for the PP test is even more pronounced than for the DF test in the previous section. It is also seen that the t_2 test indicates an extraordinary robustness in power, and even though the t_2 test is based on auxiliary regression equations, it handles smooth or abrupt changes in mean, dynamics, and the time trends comfortably.

Table 2.9 (High-High) Empirical power of the t_2 test and the corresponding Phillips-Perron t-type of test.

$T = 250$		$\pi_{23} = 0.005$				$\pi_{23} = 0.0075$				$\pi_{23} = 0.01$			
		t_2		PP		t_2		PP		t_2		PP	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
	0.10	0.07	0.07	1.00	1.00	0.08	0.08	1.00	1.00	0.09	0.09	1.00	0.98
0.01	0.15	0.08	0.09	0.94	0.87	0.09	0.10	0.82	0.70	0.11	0.11	1.00	0.46
	0.20	0.11	0.11	0.26	0.07	0.13	0.13	0.07	0.01	0.14	0.15	0.73	0.00
	0.10	0.58	0.76	0.14	0.00	0.69	0.81	0.01	0.00	0.77	0.85	0.00	0.00
0.10	0.15	0.76	0.85	0.00	0.00	0.82	0.87	0.00	0.00	0.86	0.89	0.00	0.00
	0.20	0.82	0.83	0.00	0.00	0.83	0.84	0.00	0.00	0.87	0.85	0.00	0.00
	0.10	0.72	0.87	0.00	0.00	0.81	0.88	0.00	0.00	0.86	0.90	0.00	0.00
1.00	0.15	0.86	0.90	0.00	0.00	0.84	0.88	0.00	0.00	0.84	0.86	0.00	0.00
	0.20	0.84	0.70	0.00	0.00	0.80	0.59	0.00	0.00	0.71	0.49	0.00	0.00

Note: The nominal size of the tests is 5%, and the results are based on 10 000 replications.

2.6 Concluding remarks

In this chapter we propose a new model that accommodate a smooth change in mean, trend, and dynamics. The model is a generalization of the LSTAR(p) model introduced in Lin and Teräsvirta (1994) because a nonlinear trend is added. In particular this model should be an appealing alternative when testing the null hypothesis of a random walk with drift.

Several unit root tests are derived in both the LSTAR(p) and the LSTART(p) models. The tests are enhanced by a third-order Taylor approximation, and inference about unit roots is based on the LS estimators obtained from the auxiliary testing equations. The distributional asymptotic theory for the LS estimators is non-standard because of the unit root assumption. In order to find these asymptotic distributions, we generalize many theoretical results that are derived in the econometric unit root literature. In comparison to the related approaches by Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), we note that the nonlinear models under the alternative are similar, but it should be pointed out that their testing methodology is different and no analytical limiting distributions for their unit root tests are provided.

Finite-sample properties are investigated and our tests indicate modest size distortions. The empirical power is compared to the power of the conventional unit root tests by Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988) when the DGP is an LSTAR and LSTART model, respectively. It turns out that the canonical tests only have satisfactorily power when the LSTAR and LSTART models appear linear, and under these circumstances our tests are penalized due to the many explanatory variables. Assuming more nonlinear LSTAR and LSTART models, the robustness in power of our tests in a wide range of parameter values, due to a third-order Taylor approximation, is rather extraordinary and equal unity in most of the situations. The opposite holds for the conventional tests, and the power is close to or equal to zero.

Appendix A

Mathematical Appendix

The proofs given in this Appendix are only for $m = 1$ (i.e. the LSTAR(p) models). The proofs for $m = 2, 3$ are similar and therefore omitted.

Proof of Lemma 1. The proof is straightforward and follows by induction. ■

Proof of Theorem 2.1. Write

$$\gamma_{11}\left(\hat{\psi}_1 - \psi_1\right) = \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{D}}_1 \\ \tilde{\mathbf{E}}_1 \end{bmatrix},$$

where

$$\begin{aligned}
\tilde{\mathbf{A}}_1 &= [\tilde{a}_{ij}]_{4\times 4}, \text{ and } \tilde{a}_{ij} = T^{-(i+j-1)} \sum_{t=1}^{T} t^{(i+j-2)}, \\
\tilde{\mathbf{B}}_1 &= \left[\tilde{b}_{ij}\right]_{4\times 4}, \text{ and } \tilde{b}_{ij} = T^{-(i+j-1/2)} \sum_{t=1}^{T} t^{(i+j-2)} y_{t-1}, \\
\tilde{\mathbf{C}}_1 &= [\tilde{c}_{ij}]_{4\times 4}, \text{ and } \tilde{c}_{ij} = T^{-(i+j)} \sum_{t=1}^{T} t^{(i+j-2)} y_{t-1}^2, \\
\tilde{\mathbf{D}}_1 &= [\tilde{d}_i]_{4\times 1}, \text{ and } \tilde{d}_i = T^{-(i-1/2)} \sum_{t=1}^{T} t^{i-1} u_t, \\
\tilde{\mathbf{E}}_1 &= [\tilde{e}_i]_{4\times 1}, \text{ and } \tilde{e}_i = T^{-i} \sum_{t=1}^{T} t^{i-1} y_{t-1} u_t.
\end{aligned}$$

By using Lemma 1 in Chapter 1 we conclude that the following expressions

converge jointly for all i and j:

$$\tilde{a}_{ij} \xrightarrow{d} a_{ij}, \quad \tilde{b}_{ij} \xrightarrow{d} \sigma_u b_{ij}, \quad \tilde{c}_{ij} \xrightarrow{d} \sigma_u^2 c_{ij},$$
$$\tilde{d}_i \xrightarrow{d} \sigma_u d_i, \quad \tilde{e}_i \xrightarrow{d} \sigma_u^2 e_i.$$

Thus, (2.24) holds.

Furthermore, $plim_{T\to\infty}(\hat{\psi}_1 - \psi_1) = \mathbf{0}$ holds because the element with the slowest rate of convergence in $\hat{\psi}_1$ equals $\mathcal{O}_p(T^{-1/2})$. ■

Proof of Corollary 2.2. It follows from Theorem 2.1 and the Slutsky Theorem that

$$\begin{aligned} \psi_1 &\equiv T(\hat{\varphi}_{10} - 1) \\ &= \mathbf{r}_1 \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{D}}_1 \\ \tilde{\mathbf{E}}_1 \end{bmatrix} \xrightarrow{d} \mathbf{r}_1 \mathbf{\Psi}_1^{-1} \mathbf{\Pi}_1. \end{aligned}$$

Define the matrix $\mathbf{S}_1 = diag\left\{ \begin{array}{cccccccc} 1 & 1 & 1 & 1 & \sigma_u & \sigma_u & \sigma_u & \sigma_u \end{array} \right\}$. Then

$$\begin{aligned} \mathbf{r}_1 \mathbf{\Psi}_1^{-1} \Pi_1 &= \mathbf{r}_1 \left(\mathbf{S}_1^{-1} \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1} \mathbf{S}_1^{-1} \right) \left(\sigma_u \mathbf{S}_1 \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix} \right) \\ &= \mathbf{r}_1 \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix}, \end{aligned}$$

because $\sigma_u \mathbf{r}_1 \mathbf{S}_1^{-1} = \mathbf{r}_1$. By Theorem 2.1, (2.25) holds for $m = 1$. ■

Proof of Corollary 2.3. We first note that

$$\begin{aligned} T^2 \hat{\sigma}^2_{\hat{\varphi}_{10}} &= S_T^2 \mathbf{r}_1 \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}^{-1} \mathbf{r}_1' \\ &\xrightarrow{d} \mathbf{r}_1 \tilde{\mathbf{\Psi}}_1^{-1} \mathbf{r}_1', \end{aligned}$$

where $S_T^2 = \left(y_t - \left(\mathbf{s}_{1t}', (y_{t-1}\mathbf{s}_{1t})' \right) \hat{\psi}_1 \right)^2 / (T-8)$. From Theorem 2.1, $plim_{T\to\infty} S_T^2 = \sigma^2$ holds. By Corollary 2.2 and the Slutsky Theorem we conclude that

$$\begin{aligned} t_1 &\equiv (\hat{\varphi}_{10} - 1)/\hat{\sigma}_{\hat{\varphi}_{10}} \\ &= \frac{T(\hat{\varphi}_{10} - 1)}{\left(T^2 \hat{\sigma}_{\hat{\varphi}_{m0}}\right)^{1/2}} \xrightarrow{d} \frac{\mathbf{r}_1 \tilde{\psi}_1^{-1} \tilde{\Pi}_1}{\left(\mathbf{r}_1 \tilde{\psi}_1^{-1} \mathbf{r}_1'\right)^{1/2}}. \end{aligned}$$

Thus, (2.26) holds for $m = 1$. ■

To prove Theorem 2.4, we introduce the following lemma.

Lemma 2 *If $\{u_t\}_{t=1}^{\infty}$ and $\{v_t\}_{t=1}^{\infty}$ satisfy Assumptions 1 and 2, respectively, and $\xi_t = \xi_{t-1} + v_t$ with $P(\xi_0 = 0) = 1$, then as $T \to \infty$*

$$T^{-(p+q/2+1)} \sum_{t=1}^{T} t^p \xi_{t-1}^q \xrightarrow{d} \lambda^q \int_0^1 r^p W(r)^q dr, \tag{A.1}$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p v_t v_{t-h} \xrightarrow{a.s.} \gamma_h/(p+1), \tag{A.2}$$

$$T^{-(v+1/2)} \sum_{t=1}^{T} t^v v_{t-h} \xrightarrow{d} \lambda W(1) - v\lambda \int_0^1 r^{v-1} W(r) dr, \tag{A.3}$$

$$T^{-(v+1/2)} \sum_{t=1}^{T} t^v u_t \xrightarrow{d} \sigma_u W(1) - v\sigma_u \int_0^1 r^{v-1} W(r) dr, \tag{A.4}$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t \xrightarrow{d}$$

$$0.5\lambda\sigma_u \left(W(1)^2 - p \int_0^1 r^{p-1} W(r)^2 dr - \frac{1}{(p+1)} \right), \tag{A.5}$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} v_{t-h} \xrightarrow{d}$$

$$(i) \begin{cases} 0.5 \left(\lambda^2 W(1)^2 - p\lambda^2 \int_0^1 r^{p-1} W(r)^2 dr - \dfrac{\gamma_0}{(p+1)} \right), \\ \textit{for } h = 0, \end{cases} \tag{A.6}$$

$$(ii) \begin{cases} 0.5 \left(\lambda^2 W(1)^2 - p\lambda^2 \int_0^1 W(r)^2 r^{p-1} dr - \dfrac{\gamma_0}{(p+1)} \right) + \dfrac{\sum_{s=0}^{h-1} \gamma_s}{p+1}, \\ \textit{for } h > 0, \end{cases}$$

$$T^{-(p+1/2)} \sum_{t=1}^{T} t^p v_{t-h} u_t \xrightarrow{d} N\left(0, \frac{\gamma_0 \sigma_u^2}{2p+1}\right), \tag{A.7}$$

where $\lambda = c(1)\sigma_u$, $\gamma_h = \sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h}$, $h, v, p \geq 0$, and $q \geq 1$.

Note that Lemma 2 gives more general results than needed to prove Theorem 2.4. In fact, it enables us to derive the limiting distribution for the NADF and NPADF tests in the LSTAR(p), LSTART(p), and the LSTD(p) models based on any order of approximation of a logistic transition function that accommodates an arbitrary number of transitions. Lemma 2 also generalizes many of the results derived in Hamilton (1994).

Proof of (A.1) in Lemma 2. Define the following cadlag function on $D[0,1]$,

$$W_t(r,\omega) = \frac{1}{\sigma_u\sqrt{T}}\xi_{[Tr]} = \frac{1}{\sigma_u\sqrt{T}}\xi_{t-1}, \qquad r \in [\tfrac{t-1}{T}, \tfrac{t}{T}), \qquad t = 1, ..., T,$$

where $\xi_t = \xi_{t-1} + v_t$. Using the Beveridge and Nelson (1981) decomposition we conclude that $W_t \xrightarrow{d} c(1)W$, where W is a standard Brownian motion on $[0,1]$. We can rely upon the proof for (A.1) in Lemma 1 in Chapter 1, where we replace σ with $c(1)\sigma_u$. ■

Proof of (A.2) in Lemma 2. By Assumption 2

$$T^{-1}\sum_{t=1}^{T} v_t v_{t-h} \xrightarrow{a.s.} E\left[v_t v_{t-h}\right] = \gamma_h = \sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h},$$

where $\sigma_u^2 = E\left[u_t^2\right]$, see Phillips and Solo (1992) for a proof. Furthermore,

$$\begin{aligned}
& T^{-(p+1)}\sum_{t=1}^{T} t^p v_t v_{t-h} \\
= \; & T^{-1}\sum_{t=1}^{T}(t/T)^p\left(v_t v_{t-h} - Ev_t v_{t-h} + Ev_t v_{t-h}\right) \\
= \; & T^{-1}\sum_{t=1}^{T}(t/T)^p(v_t v_{t-h} - Ev_t v_{t-h}) + T^{-1}\sum_{t=1}^{T}(t/T)^p E[v_t v_{t-h}] \\
& \xrightarrow{a.s.} \sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h}/(p+1),
\end{aligned}$$

because $\left|T^{-1}\sum_{t=1}^{T}(t/T)^p(v_t v_{t-h} - Ev_t v_{t-h})\right| \leq T^{-1}\sum_{t=1}^{T}\left|(v_t v_{t-h} - Ev_t v_{t-h})\right| \xrightarrow{a.s.}$ 0, and

$$\begin{aligned}
\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}(t/T)^p E[v_t v_{t-h}] &= \lim_{T\to\infty}\left\{(\sigma_u^2\sum_{j=0}^{\infty} c_j c_{j+h})T^{-1}\sum_{t=1}^{T}(t/T)^p\right\} \\
&= (\sigma_u^2\sum_{j=0}^{\infty} c_j c_{j+h})/(p+1) = \gamma_h/(p+1)
\end{aligned}$$

hold. ■

Proof of (A.3) in Lemma 2. This proof follows from (A.3) in Lemma 1 in Chapter 1, because

$$\begin{aligned} & T^{-(v+1/2)} \sum_{t=1}^{T} t^{v} v_{t-h} \\ = \; & T^{-1/2} \sum_{t=1}^{T} v_{t-h} - T^{-(v+1/2)} v \sum_{t=1}^{T} t^{v-1} \xi_{t-h-1} + O_p(T^{-1}) \\ & \xrightarrow{d} \sigma_u c(1) W(1) - \sigma_u c(1) v \int_0^1 r^{v-1} \; W(r) dr, \end{aligned}$$

holds for $h \geq 0$, and it is assumed that $u_{-h}, ..., u_0$ are additional starting values or random variables with finite second moments. ■

Proof of (A.4) and (A.5) in Lemma 2. These are special cases of (A.3) and (A.4), respectively, in Lemma 1 in Chapter 1. ■

Proof of (A.6)(i) in Lemma 2. Consider first the case $h = 0$, we then have that

$$\begin{aligned} & T^{-(p+1)} \sum_{t=1}^{T} t^{p} \xi_{t-1} v_t \\ & \xrightarrow{d} 0.5 \left(\lambda^2 W(1)^2 - p\lambda^2 \int_0^1 \; W(r)^2 r^{p-1} dr \; - \frac{\gamma_0}{(p+1)} \right) \end{aligned} \tag{A.8}$$

follows from (A.4) in Lemma 1 in Chapter 1, where we replace σ^2 and $\bar{\sigma}_u^2$ with λ^2 and γ_0, respectively. ■

Proof of (A.6)(ii) in Lemma 2. For $h > 0$ notice that the following recursive relationship,

$$\begin{aligned} \xi_{t-1} v_{t-h} &= \left(\xi_{t-(h+1)} + \underbrace{v_{t-h} + v_{t-h+1} + ... + v_{t-1}}_{h \text{ - terms}} \right) v_{t-h} \\ &= \xi_{t-(h+1)} v_{t-h} + v_{t-h}^2 + \sum_{s=1}^{j-1} v_{t-h} v_{t-h+s}, \end{aligned}$$

holds, and implies

$$
\begin{aligned}
& T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} v_{t-h} \\
= \; & T^{-1} \sum_{t=1}^{T} (t/T)^p \xi_{t-(h+1)} v_{t-h} + T^{-1} \sum_{t=1}^{T} (t/T)^p v_{t-h}^2 \\
& + T^{-1} \sum_{t=1}^{T} (t/T)^p \sum_{s=1}^{j-1} v_{t-h} v_{t-h+s} \\
& \xrightarrow{d} 0.5 \left(\lambda^2 W(1)^2 - p\lambda^2 \int_0^1 W(r)^2 r^{p-1} dr - \frac{\gamma_0}{(p+1)} \right) \\
& + \frac{\gamma_0}{(p+1)} + \frac{\sum_{s=1}^{h-1} \gamma_s}{(p+1)},
\end{aligned}
$$

where the convergence follows from the results in (A.2) and (A.8) ■

Proof of (A.7) in Lemma 2. Write

$$
\begin{aligned}
& T^{-1} \sum_{t=1}^{T} (t/T)^{2p} v_{t-h}^2 u_t^2 \\
= \; & T^{-1} \sum_{t=1}^{T} (t/T)^{2p} (v_{t-h}^2 u_t^2 - E v_{t-h}^2 u_t^2) + T^{-1} \sum_{t=1}^{T} (t/T)^{2p} E v_{t-h}^2 u_t^2 \\
& \xrightarrow{a.s.} \frac{\gamma_0 \sigma_u^2}{(2p+1)}
\end{aligned}
$$

because $T^{-1} \sum_{t=1}^{T} (t/T)^{2p} (v_{t-h}^2 u_t^2 - E v_{t-h}^2 u_t^2) \xrightarrow{a.s.} 0$ and notice that $E v_{t-h}^2 u_t^2 = \gamma_0 \sigma_u^2 = \sigma_u^4 \sum_{s=0}^{\infty} c_s^2 < \infty$ combined with $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} (t/T)^{2p} = 1/(2p+1)$ gives the convergence results. Using a Central Limit Theorem (CLT) for Martingale Difference Sequences (MDS) ($\{(t/T)^p v_{t-h} u_t\}$ defines a MDS) we conclude that

$$
T^{-(p+1/2)} \sum_{t=1}^{T} t^p v_{t-h} u_t \xrightarrow{d} N\left(0, \frac{\gamma_0 \sigma_u^2}{2p+1}\right).
$$

■

Proof of Theorem 2.4. Write

$$
\gamma_{12}^a \left(\hat{\psi}_1^a - \psi_1^a \right) = \begin{bmatrix} \tilde{\mathbf{F}}_1 & \tilde{\mathbf{G}}_1 \\ \tilde{\mathbf{G}}_1' & \tilde{\mathbf{H}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{K}}_1 \\ \tilde{\mathbf{L}}_1 \end{bmatrix},
$$

where

$$\begin{aligned}
\tilde{\mathbf{F}}_1 &= \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}_{8\times 8}, \\
\tilde{\mathbf{G}}_1 &= \mathbf{0}_{8\times 4(p-1)} \text{ under } H_{01}^{a,aux}, \\
\tilde{\mathbf{H}}_1 &= \left[\mathbf{T}_{11}^{-1}\left(\sum\nolimits_{t=1}^{T}(\mathbf{s}_{1t}\mathbf{s}_{1t}')\otimes\left[\tilde{h}_{ij}\right]_{(p-1)\times(p-1)}\right)T_{11}^{-1}\right]_{4(p-1)\times 4(p-1)}, \text{ with} \\
\tilde{h}_{ij} &= v_{t-i}v_{t-j} \quad \text{for } i,j\geq 1, \\
\mathbf{T}_{11} &= diag\left\{\left(\left[\begin{array}{cccc} T^{1/2} & T^{3/2} & T^{5/2} & T^{7/2}\end{array}\right]'\otimes\mathbf{1}\right)'_{1\times 4(p-1)}\right\}, \\
\tilde{\mathbf{K}}_1 &= \begin{bmatrix} \tilde{\mathbf{D}}_1 \\ \tilde{\mathbf{E}}_1 \end{bmatrix}_{8\times 1}, \\
\tilde{\mathbf{L}}_1 &= \left[\mathbf{T}_{11}^{-1}\left(\sum\nolimits_{t=1}^{T}\mathbf{s}_{1t}\otimes[\tilde{l}_i]_{(p-1)\times 1}\right)\right]_{4(p-1)\times 1}, \text{ with } \tilde{l}_i = v_{t-i}u_t \text{ for } i\geq 1.
\end{aligned}$$

Noticing that $\tilde{\mathbf{L}}_1$ defines a vector MDS with covariance matrix $\sigma_u^2\mathbf{A}_1\otimes\mathbf{\Sigma}_1$ and obeys a CLT for MDS's, applying Lemma 1 in Chapter 1, and Lemma 2, gives that the following expressions converge jointly under the null hypothesis:

$$\begin{aligned}
&\tilde{\mathbf{G}}_1 \xrightarrow{p} \mathbf{0}, \quad \tilde{\mathbf{F}}_1 \xrightarrow{d} \mathbf{\Psi}_1^a, \quad \tilde{\mathbf{H}}_1 \xrightarrow{d} \mathbf{A}_1\otimes\mathbf{\Sigma}_1, \\
&\tilde{\mathbf{K}}_1 \xrightarrow{d} \mathbf{\Pi}_1^a, \quad \tilde{\mathbf{L}}_1 \xrightarrow{d} \mathbf{R}
\end{aligned}$$

where

$$\mathbf{\Psi}_1^a = \begin{bmatrix} \mathbf{A}_1 & \lambda\mathbf{B}_1 \\ \lambda\mathbf{B}_1' & \lambda^2\mathbf{C}_1 \end{bmatrix}, \quad \mathbf{\Pi}_1^a = \begin{bmatrix} \sigma\mathbf{D}_1 \\ \sigma\lambda\mathbf{E}_1 \end{bmatrix}, \quad \mathbf{R}\sim N(\mathbf{0},\sigma_u^2\mathbf{A}_1\otimes\mathbf{\Sigma}_1).$$

By using the Slutsky Theorem,

$$\gamma_{12}^a\left(\hat{\psi}_1^a-\psi_1^a\right) \xrightarrow{d} \begin{bmatrix} \mathbf{\Psi}_1^a & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1\otimes\mathbf{\Sigma}_1 \end{bmatrix}^{-1}\begin{bmatrix} \Pi_1^a \\ \mathbf{R} \end{bmatrix} = \begin{bmatrix} (\mathbf{\Psi}_1^a)^{-1}\mathbf{\Pi}_1^a \\ (\mathbf{A}_1\otimes\mathbf{\Sigma}_1)^{-1}\mathbf{R} \end{bmatrix}.$$

Thus, $\mathbf{Q}_1^a = \mathbf{V}_1^{-1}\mathbf{R}\sim N(\mathbf{0},\sigma_u^2\mathbf{V}_1^{-1})$.

Furthermore, the fact that $\hat{\psi}_1^a$ is consistent is an immediate consequence of (2.27). ■

Proof of Theorem 2.5. This is similar to the proof of Theorem 2.4 and is therefore omitted. ■

Proof of Corollary 2.6. First, in the case ψ_1^a define $\mathbf{r}_1^* = \left[\begin{array}{cc}\mathbf{r}_1 & \mathbf{0}'\end{array}\right]$, where $\mathbf{0}$ is a $4(p-1)\times 1$ vector of zeros. It follows from Theorem 2.4 that

$$\psi_1^a \equiv T(\hat{\rho}_{10}^a - 1) = \mathbf{r}_1^*\gamma_{12}^a(\hat{\psi}_1^a - \psi_1^a) \xrightarrow{d} \mathbf{r}_1^*(\mathbf{\Psi}_1^a)^{-1}\mathbf{\Pi}_1^a = \mathbf{r}_1\mathbf{P}_1^a. \tag{A.9}$$

Furthermore, define $\mathbf{S}_3 = diag\left\{\begin{array}{cccccccc} 1 & 1 & 1 & 1 & \lambda & \lambda & \lambda & \lambda \end{array}\right\}$. Then,

$$\begin{aligned}
\mathbf{r}_1\mathbf{P}_m^a &= \mathbf{r}_1\left(\mathbf{S}_3^{-1}\begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1}\mathbf{S}_3^{-1}\right)\left(\sigma_u\mathbf{S}_3\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix}\right) \\
&= \frac{\sigma_u}{\lambda}\mathbf{r}_1\begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix}.
\end{aligned}$$

and pre-multiplying with λ/σ_u in (A.9) ensures that (2.29) holds for $m = 1$.

Second, in the case ψ_1^{pa}, define $\tilde{\mathbf{r}}_1 = \begin{bmatrix} \mathbf{r}_1 & \mathbf{0}' \end{bmatrix}$, where $\mathbf{0}$ is a $(p-1)\times 1$ vector of zeros. It follows from Theorem 2.5 that

$$\psi_1^{pa} \equiv T(\hat{\rho}_{10}^{pa} - 1) = \tilde{\mathbf{r}}_1\gamma_{12}^{pa}(\hat{\psi}_1^{pa} - \psi_1^{pa}) \xrightarrow{d} \tilde{\mathbf{r}}_1\left(\mathbf{\Psi}_1^{pa}\right)^{-1}\mathbf{\Pi}_1^{pa} = \mathbf{r}_1\mathbf{P}_m^a,$$

and that (2.30) now holds follows from the first part of the proof. ■

Proof of Corollary 2.7. We only give the proof for (2.31), since the proof for partial augmentation is similar. Hence,

$$\begin{aligned}
T^2\hat{\sigma}^2_{\hat{\rho}_{10}^a} &= (S_T^a)^2\mathbf{r}_1^*\begin{bmatrix} \tilde{\mathbf{F}}_1 & \tilde{\mathbf{G}}_1 \\ \tilde{\mathbf{G}}_1' & \tilde{\mathbf{H}}_1 \end{bmatrix}^{-1}\mathbf{r}_1^{*\prime} \\
&\xrightarrow{d} \sigma_u^2\mathbf{r}_1^*\begin{bmatrix} \mathbf{\Psi}_1^a & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1\otimes\mathbf{\Sigma}_1 \end{bmatrix}^{-1}\mathbf{r}_1^{*\prime} \\
&= \sigma_u^2\mathbf{r}_1\left(\mathbf{\Psi}_1^a\right)^{-1}\mathbf{r}_1' \\
&= \sigma_u^2\mathbf{r}_1\mathbf{S}_3^{-1}\tilde{\mathbf{\Psi}}_1^{-1}\mathbf{S}_3^{-1}\mathbf{r}_1' \\
&= \sigma_u^2/\lambda^2\mathbf{r}_1\tilde{\mathbf{\Psi}}_1^{-1}\mathbf{r}_1',
\end{aligned}$$

where $(S_T^a)^2 = \left(y_t - \left(\mathbf{s}_{1t}', (y_{t-1}\mathbf{s}_{1t})', (\mathbf{s}_{1t}\otimes\Delta y_t)'\right)\hat{\psi}_1^a\right)^2/(T - 8 - 4(p-1))$. It follows from Theorem 2.4 that $(S_T^a)^2 \xrightarrow{p} \sigma_u^2$. Using Corollary 2.6 and the Slutsky Theorem we conclude that

$$t_1^a \equiv (\hat{\rho}_{10}^a - 1)/\hat{\sigma}_{\hat{\rho}_{10}^a} = \frac{T(\hat{\rho}_{10}^a - 1)}{\left(T^2\hat{\sigma}^2_{\hat{\rho}_{10}^a}\right)^{1/2}} \xrightarrow{d} \frac{\mathbf{r}_1\tilde{\mathbf{\Psi}}_1^{-1}\tilde{\mathbf{\Pi}}_1}{\left(\mathbf{r}_1\tilde{\mathbf{\Psi}}_1^{-1}\mathbf{r}_1'\right)^{1/2}}.$$

■

Appendix B

Additional Tables

Table B.1 Critical values for the NPADF tests $c^*\psi_m^{pa}$ and t_m^{pa} in Corollaries 2.6 and 2.7 with $p = 3$.

	$c^*\psi_1^{pa}$			t_1^{pa}		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-451.27	-230.45	-170.90	-3.99	-3.21	-2.81
100	-211.43	-143.54	-115.99	-3.89	-3.15	-2.77
250	-172.32	-122.21	-98.90	-3.83	-3.11	-2.71
500	-163.55	-115.41	-93.87	-3.80	-3.08	-2.70
1000	-160.35	-112.81	-91.20	-3.79	-3.06	-2.67
∞	-156.46	-109.5	-89.89	-3.77	-3.05	-2.66
	$c^*\psi_2^{pa}$			t_2^{pa}		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-655.89	-302.44	-219.37	-3.93	-3.16	-2.75
100	-247.81	-171.04	-138.84	-3.83	-3.09	-2.71
250	-197.13	-141.11	-115.12	-3.74	-3.02	-2.63
500	-188.72	-133.50	-109.12	-3.71	-3.00	-2.61
1000	-182.92	-130.21	-105.79	-3.69	-2.97	-2.59
∞	-175.60	-126.55	-103.15	-3.66	-2.95	-2.57
	$c^*\psi_3^{pa}$			t_3^{pa}		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-190.16	-98.82	-74.52	-5.32	-4.63	-4.30
100	-79.46	-56.64	-47.75	-5.18	-4.57	-4.27
250	-58.06	-45.00	-39.09	-5.08	-4.53	-4.25
500	-53.48	-42.22	-37.06	-5.07	-4.52	-4.25
1000	-51.45	-41.11	-36.03	-5.06	-4.52	-4.24
∞	-48.70	-39.87	-34.99	-5.06	-4.51	-4.24

Notes: The probability shown at the head of each column is the area in the left-hand tail. The results are based on 100 000 replications when the true values of the nuisance parameters are zero.

Table B.2 Critical values for the NPADF tests $c^*\psi_m^{pa}$ and t_m^{pa} in Corollaries 2.6 and 2.7 with $p = 4$.

	$c^*\psi_1^{pa}$			t_1^{pa}		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-823.27	-663.55	-364.47	-4.09	-3.32	-2.92
100	-322.00	-197.03	-151.67	-3.97	-3.23	-2.84
250	-190.65	-133.64	-107.55	-3.89	-3.15	-2.76
500	-173.87	-121.52	-98.11	-3.83	-3.12	-2.73
1000	-163.29	-115.74	-93.81	-3.80	-3.09	-2.69
∞	-156.46	-109.5	-89.89	-3.77	-3.05	-2.66
	$c^*\psi_2^{pa}$			t_2^{pa}		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-1553.21	-722.89	-488.81	-4.05	-3.28	-2.88
100	-419.25	-247.02	-189.14	-3.91	-3.17	-2.79
250	-222.47	-156.82	-127.78	-3.81	-3.09	-2.70
500	-198.86	-141.57	-115.27	-3.75	-3.03	-2.65
1000	-188.85	-134.74	-109.63	-3.71	-2.99	-2.62
∞	-175.60	-126.55	-103.15	-3.66	-2.95	-2.57
	$c^*\psi_3^{pa}$			t_3^{pa}		
T	0.01	0.05	0.10	0.01	0.05	0.10
50	-1462.068	-332.69	-184.10	-5.34	-4.65	-4.32
100	-136.38	-80.78	-63.83	-5.17	-4.57	-4.27
250	-66.20	-49.74	-42.81	-5.09	-4.54	-4.26
500	-56.49	-44.02	-38.43	-5.06	-4.52	-4.25
1000	-52.79	-41.75	-36.67	-5.06	-4.51	-4.24
∞	-48.70	-39.87	-34.99	-5.06	-4.51	-4.24

Notes: The probability shown at the head of the column is the area in the left-hand tail. The results are based on 100 000 replications when the true values of the nuisance parameters are zero.

Bibliography

BANERJEE, A., R. L. LUMSDAINE, AND H. J. STOCK (1992): "Recursive and sequential tests of the unit-root and trend break hypotheses: Theory and international evidence," *Journal of Business and Economic Statistics*, 10, 271–287.

BEC, F., M. B. SALEM, AND M. CARRASCO (2002): "The purchasing power parity puzzle: Evidence from a LSTAR specification," Universit de Paris I, Working paper.

BEVERIDGE, S., AND C. R. NELSON (1981): "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle," *Journal of Monetary Economics*, 7, 151–174.

CHAN, K. S., AND H. TONG (1986): "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–191.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series," *Journal of the American Statistical Association*, 74, 427–431.

EKLUND, B. (2003a): "Testing the unit root hypothesis against the logistic smooth transition autoregressive model," Stockholm School of Economics, Working paper.

——— (2003b): "A nonlinear alternative to the unit root hypothesis," Stockholm School of Economics, Working paper.

ENDERS, W., AND C. GRANGER (1998): "Unit root tests and asymmetric adjustment with an example using the term structure of interest rates," *Journal of Business and Economic Statistics*, 16, 304–311.

HALL, A. (1989): "Testing for a unit root in the presence of moving average errors," *Biometrika*, 76, 49–56.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, New Jersey.

HARVEY, D. I., AND T. C. MILLS (2002): "Unit Roots and Double Smooth Transitions," *Journal of Applied Statistics*, 29, 675–683.

KAPETANIOS, G., Y. SHIN, AND A. SNELL (2003): "Testing for a unit root in the nonlinear STAR framework," *Journal of Econometrics*, 112, 359–379.

LANNE, M., AND H. LÜTKEPOHL (2002): "Unit root test for time series with level shifts: A comparison of different proposals," *Economic Letters*, 75, 109–114.

LANNE, M., H. LÜTKEPOHL, AND P. SAIKKONEN (2003): "Test procedures for unit roots in time series with level shifts at unknown time," *Oxford Bulletin of Economics and Statistics*, 65, 91–115.

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

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

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

OULIARIS, S., J. PARK, AND P. C. B. PHILLIPS (1989): "Testing for a unit root in the presence of a maintained trend," *Kluwer*, pp. 7–28.

PERRON, P. (1989): "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 57, 1361–1401.

——— (1990): "Testing for a unit root in a time series with a changing mean," *Journal of Business and Economic Statistics*, 8, 153–162.

PERRON, P., AND T. J. VOGELSANG (1992): "Nonstationarity and level shifts with an application to purchasing power parity," *Journal of Business and Economic Statistics*, 10, 301–320.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

SAIKKONEN, P., AND H. LÜTKEPOHL (2002): "Testing for a unit root in a time series with level shift at unknown time," *Econometric Theory*, 18, 313–348.

SCHWERT, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–159.

SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990): "Inference in Linear Time Series with Some Unit Roots," *Econometrica*, 58, 113–144.

TONG, H. (1990): *Non-linear time series.* Oxford Science publications.

ZIVOT, E., AND D. W. K. ANDREWS (1992): "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.

Chapter 3

Inference for Unit Roots in a Panel Smooth Transition Autoregressive Model Where the Time Dimension is Fixed

3.1 Introduction

One can expect the traditional panel data unit root tests by Quah (1994), Harris and Tzavalis (1999), and Im, Pesaran, and Shin (2003), to have low power if the time series for the cross sections exhibit structural shifts in levels and/or trends. One explanation to this is that the authors quoted above base their unit root tests on panels where each cross section is modelled as a linear autoregressive process considered in e.g. Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988). As such, it has been pointed out by Perron (1990) that the classical univariate unit root tests are biased towards nonrejection in time series with structural changes, and "adding" up the nonlinearities in single time series into a panel framework will most likely lead to a bias towards nonrejection for the traditional panel unit root tests as well. Considering this fact in addition to that evidence of nonlinearities (such as structural breaks) in many single, say, macroeconomic time series are found, it seems that the traditional panel data unit root tests in such cases are based on too restrictive panels. In particular, this will have serious implications for applied work because the shocks to each cross section will be treated as if they have a permanent effect.

Panel data unit root tests allowing for structural breaks can for instance be found in Im and Lee (1999), Silvestre, Barrio-Castro, and Lopez-Bazo (2001), Tzavalis (2002), and Bai and Silvestre (2003). They derive unit root tests in a panel where each cross section has an abrupt structural shift in the level and/or the time trend. However, in many cases a gradual or more smooth change between two regimes seems preferable, see e.g. Ripatti and Saikkonen (2001). In this chapter we generalize the idea with an instant shift in levels by introducing a nonlinear dynamic panel accommodating a smooth cross section specific change in levels and a homogeneous smooth shift in dynamics, in which we test the null hypothesis of a common unit root.

Our test of a common unit root is based on the normalized LSDV estimator of the autoregressive coefficient in an auxiliary regression equation. The time dimension is fixed and the cross section dimension tends to infinity. The analytical limiting distribution of the test is the standard normal where the two first moments are calculated analytically. Our approach is similar to the one in Harris and Tzavalis (1999) whose results are obtained as special cases of ours. We choose therefore to compare the power of our test to the power of their tests. This gives the opportunity to demonstrate when the tests in Harris and Tzavalis (1999) actually have substantial power when in fact a nonlinear panel is considered, as well as scenarios when the traditional tests are heavily biased towards nonrejection.

The rest of the chapter is organized as follows. In Section 3.2 we present the nonlinear dynamic panel. In Section 3.3 we present the procedure for testing a unit root and derive the test statistic. Section 3.4 contains simulation experiments to examine the finite-sample properties of the test. Concluding remarks are given in Section 3.5. Thereafter an Appendix follows where proofs can be found.

3.2 The model

Consider a first-order panel smooth transition autoregressive (PSTAR(1)) model,

$$y_{it} = \pi_{i,10}+\pi_{11}y_{i,t-1}+(\pi_{i,20}+\pi_{21}y_{i,t-1})F(t;\gamma,c)+u_{it}, \quad i=1,...,n, \quad t=1,...T, \tag{3.1}$$

where the index i represents the i'th cross sectional unit, t indexes time series observations, and u_{it} is the error term. In (3.1), $F(t;\gamma,c)$ is a transition function satisfying the conditions: (i) $F(t;\gamma,c)$ is a bounded and continuous function for all t, γ, and c. (ii) $F(t;0,c)=0$. (iii) In an open interval $(-\varepsilon,\varepsilon)$, for $\varepsilon > 0$, $\partial F(t;\gamma,c)/\partial\gamma$ is non-zero and $\partial^2 F(t;\gamma,c)/\partial\gamma^2$ exists. (iv) For fixed γ and c, $F(t;\gamma,c)$ is monotonic in t. A suitable choice of a transition function in (3.1) that meets the conditions in (i)-(iv) is the logistic cumulative distribution function (after a downward shift)

$$F(t;\gamma,c) = \frac{1}{1+\exp\{-\gamma(t-c)\}} - \frac{1}{2}. \tag{3.2}$$

In (3.2), $\gamma > 0$ is a slope parameter indicating how rapid the transition is, and $c \in (0,T)$ is a location parameter around which the transition (symmetrically) takes place. Restriction $\gamma > 0$ is an identifying restriction, and implies that $F(t)$ is increasing in t. Model (3.1) with (3.2) defines the panel logistic smooth transition autoregressive model of order one, called the PLSTAR(1) model for short. For convenience we only consider a first-order polynomial of t in (3.2). For a discussion about higher-order polynomials in t, see e.g. Lin and Teräsvirta (1994), or Teräsvirta (1998).

The PLSTAR(1) model contains nonlinear heterogeneous fixed effects, nonlinear homogeneous autoregressive coefficients, and homogeneous slope and location parameters.[1] For each individual equation i, the function $F(t;\gamma,c)$ in (3.2) allows for a smooth change between regimes in intercepts and dynamics.

[1]The homogeneity assumption imposed on some of the parameters is needed for the coming testing procedure.

It is evident that the PLSTAR(1) model specification nests many panel models studied in the literature. In particular, when $\gamma \to \infty$, $F(t;\gamma,c)$ in (3.2) becomes an indicator function, i.e. $F(t;\infty,c) = -0.5$ if $t \in [0,c)$ and $F(t;\infty,c) = 0.5$ if $t \in [c,T]$, and the PLSTAR(1) model displays a panel threshold AR(1) (PTAR(1)) model with a single structural break at $t = c$. At the other end, by letting $\gamma = 0$ in (3.2) the PLSTAR(1) model collapses into a linear panel AR(1) (PAR(1)) model. It may be mentioned that González, Teräsvirta, and van Dijk (2004) recently introduced a different panel STAR model by generalizing the panel threshold model of Hansen (1999). In their model, $y_{i,t-1}$, is replaced by a vector of exogenous variables $\mathbf{x}_{it}$ so the model is not dynamic and unit roots are not an issue. Furthermore, the transition variable in the transition function (3.2) is a stochastic variable which can be an element of $\mathbf{x}_{it}$. Even there, when $\gamma \to 0$ in (3.2), the model becomes a linear homogeneous panel model with exogenous variables.

3.3 Test statistic

We now consider a test statistic for testing the hypothesis of a panel unit root in the PLSTAR(1) model. Under this hypothesis, $H_0 : \pi_{i,10} \in \mathbb{R}$ for all i, $\pi_{11} = 1$, and $\gamma = 0$, in (3.1) and (3.2).[2] It is tested against a stable PLSTAR(1) model with $\gamma > 0$. The stability conditions are given by $\pi_{11} - 0.5\pi_{21} \in (-1,1)$ and $\pi_{11} + 0.5\pi_{21} \in (-1,1)$ to rule out non-stationary or explosive trajectories. Note, however, that the PLSTAR(1) model also becomes linear by for any i setting $\pi_{i,20} = \pi_{21} = 0$ in (3.1). This shows that there is an identification problem in the PLSTAR(1) model under the null hypothesis $\gamma = 0$ because then the parameters $\pi_{i,20}$, π_{21}, and c are not identified. We circumvent this difficulty by an approximation of $F(t;\gamma,c)$, as suggested by Luukkonen, Saikkonen, and Teräsvirta (1988). An obvious candidate is the first-order Taylor expansion of $F(t;\gamma,c)$ around $\gamma = 0$. Applying this approximation to (3.2) and merging terms and reparameterizing, we obtain the following version of the PLSTAR(1) model

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \delta_i t + \phi t y_{i,t-1} + u^*_{it}, \tag{3.3}$$

where u^*_{it} is an adjusted error term such that $u^*_{it} = u_{it}$ holds under the null hypothesis, i.e. the distributional properties of the error process are preserved under the null hypothesis and are not affected by the Taylor approximation.

[2]That is, a joint test of parameter constancy (linearity) and a unit root.

The parameters α_i, ρ, δ and ϕ are all functions of the originally defined parameters such that the originally stated null hypothesis is transformed into

$$H_0^{aux}: \ \alpha_i \in \mathbb{R} \quad \text{for all } i, \quad \rho = 1, \quad \delta_i = 0, \quad \phi = 0. \tag{3.4}$$

Note that the linear models in Harris and Tzavalis (1999) are nested in the auxiliary regression (3.3), and we thereby find two plausible competing tests in their Theorems 2 and 3, which will be referred to as the HT2 and HT3 tests respectively. This will also give the opportunity to examine the expectation that traditional panel data unit root test are biased towards nonrejection under models with a shift in levels. Furthermore,The HT2 and HT3 tests are based on models letting $(\delta_i, \phi) = (0, 0)$ and $\phi = 0$, respectively, in (3.3). To proceed we impose the following assumptions on the PLSTAR(1) model.

Assumption 1 *(A1) Let $\{u_{it}\}_{i,t\in\mathbb{N}}$ be an i.i.d. sequence of random variables such that $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_u^2$ hold for all i and t. (A2) The individual effect α_i equals 0 for all i. (A3) $E(u_{it}^4) = \mu_4 < \infty$ for all i and t.*

3.3.1 Bias estimator

The panel unit root test statistic is constructed from the normalized coefficient statistic based directly on the LSDV estimator of the coefficients of the auxiliary fixed effect model (3.3). Under the null hypothesis (3.4), the deviation form of the LSDV estimator of ρ in the model (3.3) is given by

$$\widehat{\rho} - 1 = \frac{\left[\left(\sum_{i=1}^{n} W_{i,2t}\right)\left(\sum_{i=1}^{n} W_{i,4t}\right) - \left(\sum_{i=1}^{n} W_{i,3t}\right)\left(\sum_{i=1}^{n} W_{i,5t}\right)\right]}{\left[\left(\sum_{i=1}^{n} W_{i,1t}\right)\left(\sum_{i=1}^{n} W_{i,2t}\right) - \left(\sum_{i=1}^{n} W_{i,3t}\right)^2\right]}. \tag{3.5}$$

In (3.5), $W_{i,1t} = \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{y}_{i,-1}$, $W_{i,2t} = \mathbf{y}_{i,-1}'\mathbf{D}_T\,\mathbf{Q}_T\mathbf{D}_T\mathbf{y}_{i,-1}$, $W_{i,3t} = \mathbf{y}_{i,-1}'\mathbf{Q}_T\,\mathbf{D}_T\mathbf{y}_{i,-1}$, $W_{i,4t} = \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{u}_i$, $W_{i,5t} = \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{u}_i$ with vectors $\mathbf{y}_{i,-1} = (y_{i0}, ..., y_{i,T-1})'$, $\mathbf{u}_i = (u_{i1}, ..., u_{iT})'$. Furthermore, $\mathbf{Q}_T$ is the $(T \times T)$ within transformation matrix defined by $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{M}_T$ where $\mathbf{M}_T = \mathbf{X}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{X}_T'$ with $\mathbf{X}_T = (\boldsymbol{\iota}_T, \boldsymbol{\tau}_T)$ and $\boldsymbol{\iota}_T$ is the unit column vector of length T, $\boldsymbol{\tau}_T = (1, 2, ..., T)'$, and $\mathbf{D}_T = diag\{1, 2, ..., T\}$. Under the null hypothesis (3.4) the LSDV estimator in (3.5) is inconsistent for fixed T as $n \to \infty$. This result is stated in the following theorem.

Theorem 3.1 *Consider model (3.3) when (3.4) and (A1)-(A3) in Assumption 1 hold. Then, for any fixed $T > 2$, the LSDV estimator $(\widehat{\rho} - 1)$ in (3.5) satisfies*

$$\underset{n\to\infty}{plim}(\widehat{\rho} - 1) = B_1(T), \tag{3.6}$$

where

$$B_1(T) \equiv -\frac{1}{4}\frac{23T^2 - 21T - 74}{(T^2 - 2)(T + 2)}.$$

Proof. See Appendix A. ■

Theorem 3.1 states that when n tends infinity and T is fixed, the LSDV estimator $\widehat{\rho}$ in (3.3) is inconsistent under the null hypothesis (3.4). The degree of inconsistency only depends upon T, and it is order equals $\mathcal{O}(T^{-1})$. Thus, $T \to \infty$ is required for $plim_{n,T\to\infty}\widehat{\rho} = 1$ to hold. The inconsistency arises because of the elimination of the fixed effects α_i and the time trend δ_i by the $\mathbf{Q}_T$ matrix from each observation of the panel, see Nickell (1981). This makes the explanatory variables correlated with the error term as Hsiao (1986) pointed out. An interesting feature is that the bias is negative. To see this note that under (3.4), $\mathbf{y}_{i,-1} = y_{i0}\iota_T + \mathbf{C}_T\mathbf{u}_i$ holds where $\mathbf{C}_T$ is the strictly lower triangular $(T \times T)$ matrix

$$\mathbf{C}_T = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 \end{bmatrix}, \tag{3.7}$$

which implies that the inequalities

$$E\left(\mathbf{y}'_{i,-1}\mathbf{Q}_T\mathbf{u}_i\right) = -E\left(\mathbf{u}'_i\mathbf{C}'_T\mathbf{M}_T\mathbf{u}_i\right) < 0$$

and

$$E\left(\mathbf{y}'_{i,-1}\mathbf{D}_T\mathbf{Q}_T\mathbf{u}_i\right) = -E\left(\mathbf{u}'_i\mathbf{C}'_T\mathbf{D}_T\mathbf{M}_T\mathbf{u}_i\right) < 0$$

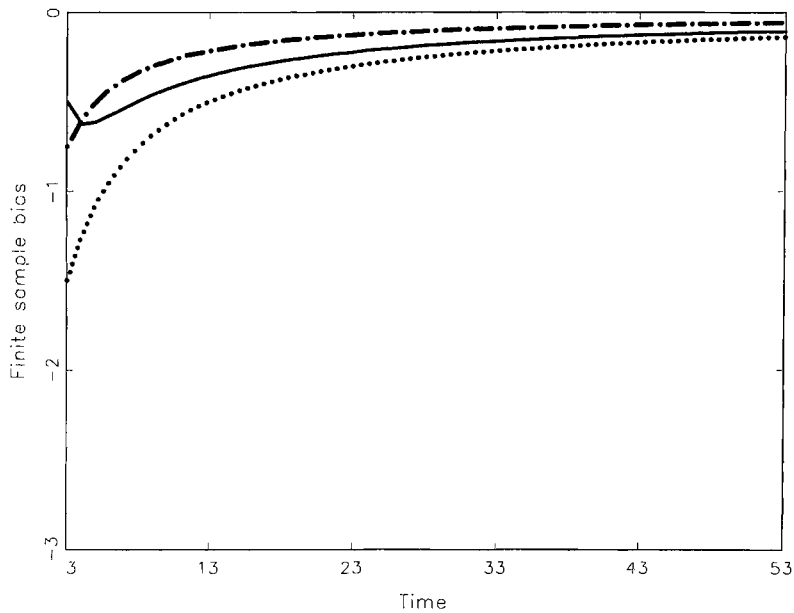
hold. By these inequalities it follows that induced correlations between the explanatory variables and the error disturbances are always negative. This feature plays a key role in yielding the negative value of $B_1(T)$. Furthermore, the bias for the HT2 and HT3 test statistics are obtainable by using (3.5), and equal

$$\underset{n\to\infty}{plim}(\widehat{\rho} - 1) = B_{HT2}(T) \equiv -3(T + 1)^{-1}, \tag{3.8}$$

$$\underset{n\to\infty}{plim}(\widehat{\rho} - 1) = B_{HT3}(T) \equiv -\frac{15}{2}(T + 2)^{-1}. \tag{3.9}$$

For $T > 3$, one can show that $|B_{HT2}(T)| < |B_1(T)| < |B_{HT3}(T)|$. In fact, for small T, $|B_1(T)| - |B_{HT3}(T)|$ could be rather substantial. This is illustrated in Figure 3.1.

Figure 3.1 The finite sample bias of the LSDV estimator in Theorem 1 (solid line), and the corresponding biases HT2 (dash-dotted line) and HT3 (dotted line) in Harris and Tzavalis (1999).



3.3.2 Asymptotic distribution

Because the bias of $(\widehat{\rho} - 1)$ is known, it is possible to derive the limiting distribution for the bias corrected normalized statistic for the model (3.3) under the null hypothesis (3.4). The result is given in the following theorem.

Theorem 3.2 *Consider model (3.3) when (3.4) and (A1)-(A3) in Assumption 1 hold. Then, for any fixed $T > 2$, the limiting distribution of the LSDV estimator $(\widehat{\rho} - 1)$ in (3.5), adjusted by $B_1(T)$ in (3.6), is given by*

$$\sqrt{n}(\widehat{\rho} - 1 - B_1(T)) \xrightarrow{d} N\left(0, \sigma^2_{\widehat{\rho}}(T, \kappa_4)\right), \tag{3.10}$$

where

$$\sigma_{\widehat{\rho}}^2(T,\kappa_4) \equiv 5\kappa_4 \frac{n_1(T)}{n_2(T)} + \frac{n_3(T)}{n_4(T)},$$

with

$$\begin{aligned} n_1(T) &= 8428\,767T^{11} - 13\,614\,689T^{10} - 120\,059\,496T^9 \\ &\quad +186\,771\,124T^8 + 721\,928\,310T^7 - 948\,544\,018T^6 \\ &\quad -2393\,879\,224T^5 + 2116\,570\,904T^4 + 5166\,454\,483T^3 \\ &\quad +615\,163\,035T^2 - 1914\,301\,704T - 461\,936\,628, \end{aligned}$$

$$n_2(T) = 512\,512\left(T^2-2\right)^4 (T+2)^3 \left(T^2-1\right)(T-2)\,T\,(T-3)^{-1},$$

$$\begin{aligned} n_3(T) &= 686\,450\,089T^{13} - 2714\,666\,460T^{12} + 5972\,242\,321T^{11} \\ &\quad +22\,845\,456\,210T^{10} - 149\,532\,661\,418T^9 - 51\,654\,581\,616T^8 \\ &\quad +893\,153\,037\,170T^7 - 96\,760\,187\,484T^6 - 2612\,622\,746\,635T^5 \\ &\quad +322\,041\,658\,116T^4 + 4127\,083\,405\,469T^3 + 994\,368\,662\,874T^2 \\ &\quad -1478\,687\,733\,396T - 374\,168\,668\,680, \end{aligned}$$

$$n_4(T) = 9225\,216\left(T^2-2\right)^4 (T+2)^3 (T-2)(T-1)(T+1)\,T,$$

and

$$\kappa_4 = \mu_4/\sigma_u^4.$$

Proof. See Appendix A. ■

Theorem 3.2 states that the test statistic defined in (3.10), corrected by $B_1(T)$ in (3.6) for the inconsistency of $\widehat{\rho}$ in (3.5), is normally distributed with mean zero and variance $\sigma_{\widehat{\rho}}^2(T,\kappa_4)$ as $n \to \infty$. The asymptotic variance $\sigma_{\widehat{\rho}}^2(T,\kappa_4)$ is a function of T and the nuisance parameter κ_4. The dependence on κ_4 can be eliminated by imposing the normality assumption on the disturbances $\{u_{it}\}$, and the simplified form of the variance $\sigma_{\widehat{\rho}}^2(T,\kappa_4)$ in (3.10) appears in the following corollary.

Corollary 3.3 *Consider model (3.3) when (3.4) and (A1)-(A3) in Assumption 1 hold. Furthermore, assume that u_{it} are normally distributed. Then, for any fixed $T > 2$, the limiting distribution of the LSDV estimator $(\widehat{\rho}-1)$ in (3.5), adjusted by $B_1(T)$ in (3.6), is given by*

$$\sqrt{n}(\widehat{\rho} - 1 - B_1(T)) \xrightarrow{d} N(0, \sigma_{\widehat{\rho}}^2(T)), \tag{3.11}$$

where

$$\sigma^2_{\widehat{\rho}}(T) \equiv \frac{n_5(T)}{n_6(T)}, \tag{3.12}$$

with

$$\begin{aligned} n_5(T) \;=\; & 52\,803\,853T^{10} - 33\,761\,490T^9 - 295\,736\,530T^8 \\ & +78\,337\,770T^7 - 438\,526\,236T^6 - 538\,473\,642T^5 \\ & +3583\,336\,934T^4 + 1400\,993\,790T^3 - 4271\,003\,921T^2 \\ & +1598\,065\,812T + 4063\,557\,132, \end{aligned}$$

and

$$n_6(T) = 709\,632\left(T^2-2\right)^4(T+2)^3(T-2).$$

Proof. See Appendix A. ■

From Corollary 3.3 we see that the asymptotic normality of $\sqrt{n}(\widehat{\rho}-1-B_1(T))$ yields a test statistic which depends only on the estimated parameter $\widehat{\rho}$ and known values of n and T. Hence, $\sqrt{n}(\widehat{\rho}-1-B_1(T))/\sigma_{\widehat{\rho}}(T)$ can be readily used for statistical inference, and the critical values of the standard normal distribution apply. Moreover, by similar manipulations we also find the results

$$\sqrt{n}(\widehat{\rho}-1-B_{HT2}(T)) \xrightarrow{d} N(0,\sigma^2_{\widehat{\rho}_{HT2}}(T)), \tag{3.13}$$

$$\sqrt{n}(\widehat{\rho}-1-B_{HT3}(T)) \xrightarrow{d} N(0,\sigma^2_{\widehat{\rho}_{HT3}}(T)), \tag{3.14}$$

where

$$\sigma^2_{\widehat{\rho}_{HT2}}(T) \equiv \frac{3}{5}\frac{\left(17T^2-20T+17\right)}{(T-1)\,(T+1)^3},$$

and

$$\sigma^2_{\widehat{\rho}_{HT3}}(T) \equiv \frac{15}{112}\frac{(193T^2-728T+1147)}{(T-2)(T+2)^3},$$

are the variances Harris and Tzavalis (1999) obtained under the normality assumption. Note that although $|B_{HT2}(T)| < |B_1(T)| < |B_{HT3}(T)|$ for $T > 3$, one can show that $\sigma^2_{\widehat{\rho}}(T) > \sigma^2_{\widehat{\rho}_{HT3}}(T) > \sigma^2_{\widehat{\rho}_{HT2}}(T)$ holds for $T > 4$. The situation is illustrated in the left-hand panel of Figure 3.2 where it is seen that $\sigma^2_{\widehat{\rho}}(T)$ is a decreasing function in T for $T \geq 4$. In addition we note that $\max_{T\in(2,\infty)}\{\sigma^2_{\widehat{\rho}}(T)\} = 2.29$ occurs at $T = 4$. The reason for the ordering between the variances is that under the auxiliary null hypothesis (3.4), the LSDV estimators $\widehat{\rho}$ and $\widehat{\phi}$ are correlated.

Note that the result in Corollary 3.3 is specific to the case that $\alpha_i = 0$ for all i. If (A2) in Assumption 1 is relaxed the limiting distribution of Corollary 3.3 is no longer invariant with respect to α_i and σ_u^2. By including cross section specific trends in the regression model (1c) in Harris and Tzavalis (1999) it is shown that their limiting distribution of the test statistic in (3.14) is invariant with respect to α_i and σ_u^2 without assuming $\alpha_i = 0$. Although our regression in (3.3) contains a time trend, the limiting distribution in (3.11) requires that $\alpha_i = 0$ for any i due to the nonlinear feature of (3.3).

Letting both n and T tend to infinity in (3.11) results in a degenerate limiting distribution Because $\sigma_{\widehat{\rho}}^2(T)$ in (3.12) is $\mathcal{O}(T^{-2})$. For this reason it is necessary to re-scale the test statistic (3.11) in Corollary 3.3. We have the following result.

Corollary 3.4 *Suppose that the conditions in Corollary 3.3 hold for model (3.3). Then, as $T \to \infty$ and $n \to \infty$*

$$\sqrt{n}T(\widehat{\rho} - 1) + \frac{23}{4}\sqrt{n} \xrightarrow{d} N\left(0, \frac{52\,803\,853}{709632}\right). \tag{3.15}$$

Proof. The proof of (3.15) follows immediately from Corollary 3.3. ■

As noted by Levin, Lin, and Chu (2002), in contrast to the case of stationary panel data, the presence of a unit root causes the fixed effects to influence the asymptotic distribution of the panel autoregressive estimator by factor $23\sqrt{n}/4$, even as both n and T become large. Also, Corollary 3.4 implies that $\widehat{\rho}$ in (3.5) converges at the rate $\sqrt{n}T$, which is higher than the convergence rate of the LSDV estimator in the stationary case. Comparing the test statistic (3.15) with (3.11), we see that the term adjusting for inconsistency of $\widehat{\rho}$ in Corollary 3.4 when $T \to \infty$ is greater than $\sqrt{n}\,|B_1(T)|$ in Corollary 3.3 when T is fixed. Furthermore, the asymptotic variance $lim_{T\to\infty} T^2\sigma_{\widehat{\rho}}^2(T) = 52\,803\,853/709632 \approx 74.41$ is always greater than $T^2\sigma_{\widehat{\rho}}^2(T)$ for any fixed T, see the right-hand panel of Figure 3.2.

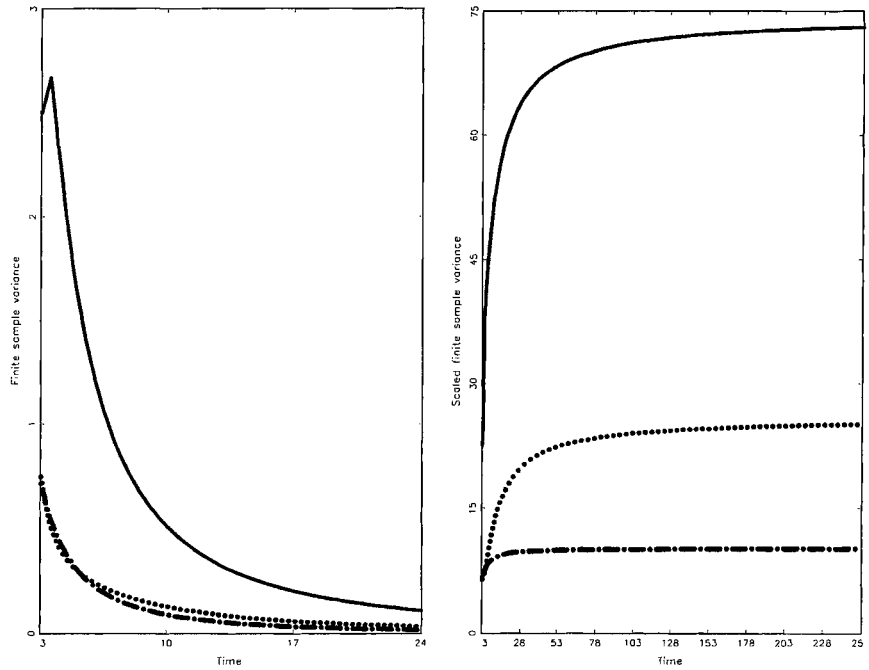
The results in Corollaries 3.3 and 3.4 may be applied to consider the consequences, as suggested by Harris and Tzavalis (1999), of assuming that T is asymptotic rather than fixed. In detail, Corollary 3.4 implies that we would use

$$\frac{T\sqrt{n}(\widehat{\rho} - 1)}{\sqrt{74.41}} + \frac{23}{4}\frac{\sqrt{n}}{\sqrt{74.41}} \underset{n,T\to\infty}{\xrightarrow{d}} N(0,1), \tag{3.16}$$

for inference when $T, n \to \infty$, whereas the true distribution for $T < \infty$ and $n \to \infty$ is

$$\frac{\sqrt{n}T(\widehat{\rho} - 1)}{\sqrt{74.41}} + \frac{23}{4}\frac{\sqrt{n}}{\sqrt{74.41}}C_1(T) \underset{n\to\infty}{\xrightarrow{d}} N(0, C_2(T)), \tag{3.17}$$

Figure 3.2 The finite sample variance $\sigma^2_{\hat{\rho}}(T)$ and the scaled finite sample variance $T^2\sigma^2_{\hat{\rho}}(T)$ (solid lines), and the corresponding variances for HT2 (dash-dotted lines) and HT3 (dotted lines) in Harris and Tzavalis (1999).



where $C_1(T) = -TB_1(T)/(23/4)$ and $C_2(T) = T^2\sigma^2_{\hat{\rho}}(T)/74.41$ such that $C_1(T), C_2(T) \in (0,1)$ holds for $4 < T < \infty$, and $\lim_{T\to\infty} C_1(T) = \lim_{T\to\infty} C_2(T) = 1$. From (3.17) it is clear that we have two possible effects if we erroneously use (3.16) in finite-samples. The first effect is the mean shift effect, $C_1 < 1$ and therefore the asymptotic distribution in (3.16) is located to the right of the finite-sample distribution in (3.17). This implies an increase in the size over the nominal level. The second effect is the variance effect, $C_2 < 1$, which implies that the asymptotic variance in (3.16) is too large so the tails of the asymptotic distribution contain excess probability mass. This leads to a decrease in the size of the standardized test statistic over the nominal level. As a conclusion, if the mean shift effect dominates the variance effect, the test will be oversized and the power is increased. The relative importance of these effects are investigated in the next section using Monte Carlo simulations.

3.3.3 Heterogeneous errors

The errors u_{it} in Assumption 1 are assumed to be i.i.d. such that $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_u^2$ holds for all i and t, but this is easily relaxed to allow for heterogeneous errors.

Assumption 2 *(B1) Let $\{u_{it}\}_{i,t\in\mathbb{N}}$ be a sequence of independently distributed random variables for all i and t with $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_i^2 < \infty$, and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sigma_i^2 = \sigma_u^2 < \infty$. (B2) The individual effect α_i equals 0 for all i. (B3) $E\left|u_{it}\right|^{4+\delta} < \infty$ holds for $\delta > 0$, and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mu_{4i} = \mu_4 < \infty$ where $Eu_{it}^4 = \mu_{4i}$.*

Assumption 2 allows us to derive the same results as in Theorems 3.1 and 3.2 and Corollaries 3.3 and 3.4 by applying the Markov Law of Large Numbers (LLN) and the Liapounov Central Limit Theorem (CLT).

3.4 Simulation experiments

In this section we conduct several Monte Carlo experiments to explore finite-sample and asymptotic properties of our test statistics defined in Corollaries 3.3 and 3.4, denoted T_1 and T_2, respectively. In addition, the properties of the HT2 and HT3 tests by Harris and Tzavalis (1999) will be investigated as well.

3.4.1 Size simulations

The aim of the first experiment is to assess the size properties of the test in Corollary 3.3. The DGP under the null hypothesis is given by

$$y_{it} = y_{i,t-1} + u_{it}, \quad i = 1, ..., n, \quad t = 1, ...T, \tag{3.18}$$

where $u_{it} \sim nid(0, 1)$ for all i and t. The empirical size and its finite-sample accuracy are reported in Table 3.1.

As can be seen from this table, the empirical distribution of the test statistic in Corollary 3.3 approximates fairly well the standard normal distribution for almost all n and T. When n is small relatively to T, there is a slight size distortion because the time dimension dominates the cross section dimension. In this case we expect the finite-sample distribution to be a less satisfactory approximation to the asymptotic distribution. However, increasing n to match the time dimensions, we see that the size discrepancy vanishes. For comparison, the bottom line in Table 3.1 reports the quantiles for the standard normal distribution.

Table 3.1 Empirical quantiles of the test statistic in Corollary 3.3.

T	n	1%	5%	10%	50%	90%	95%	99%	Size
5	5	-2.60	-1.79	-1.38	0.01	1.47	1.98	2.93	0.06
5	10	-2.44	-1.71	-1.31	0.00	1.37	1.79	2.65	0.06
5	25	-2.33	-1.63	-1.27	0.01	1.29	1.66	2.51	0.05
5	50	-2.34	-1.63	-1.27	0.01	1.31	1.70	2.43	0.05
5	100	-2.29	-1.66	-1.26	0.02	1.32	1.71	2.43	0.05
10	5	-2.55	-1.80	-1.40	-0.03	1.30	1.70	2.44	0.07
10	10	-2.50	-1.78	-1.39	-0.03	1.27	1.63	2.32	0.06
10	25	-2.48	-1.75	-1.35	-0.03	1.27	1.64	2.33	0.06
10	50	-2.41	-1.70	-1.33	-0.01	1.27	1.65	2.27	0.06
10	100	-2.43	-1.69	-1.32	-0.03	1.28	1.62	2.34	0.05
25	5	-2.90	-1.98	-1.53	-0.12	1.21	1.57	2.17	0.08
25	10	-2.64	-1.85	-1.43	-0.04	1.23	1.58	2.23	0.07
25	25	-2.49	-1.78	-1.41	-0.05	1.22	1.58	2.21	0.06
25	50	-2.44	-1.71	-1.33	-0.03	1.24	1.58	2.19	0.06
25	100	-2.43	-1.68	-1.32	-0.02	1.24	1.61	2.22	0.05
50	5	-3.10	-2.05	-1.59	-0.10	1.18	1.50	2.17	0.09
50	10	-2.69	-1.91	-1.48	-0.09	1.16	1.50	2.15	0.07
50	25	-2.67	-1.84	-1.42	-0.05	1.19	1.53	2.17	0.07
50	50	-2.50	-1.75	-1.39	-0.04	1.26	1.60	2.19	0.06
50	100	-2.46	-1.71	-1.33	-0.01	1.23	1.57	2.28	0.06
100	5	-3.11	-2.07	-1.63	-0.10	1.19	1.51	2.09	0.09
100	10	-2.80	-1.87	-1.45	-0.06	1.18	1.51	2.11	0.07
100	25	-2.72	-1.87	-1.43	-0.05	1.21	1.53	2.16	0.07
100	50	-2.52	-1.77	-1.37	-0.04	1.20	1.57	2.22	0.06
100	100	-2.48	-1.74	-1.34	-0.03	1.23	1.57	2.16	0.06
$N(0,1)$		-2.33	-1.65	-1.28	0.00	1.28	1.65	2.33	

Note: The nominal size is 5%, and the results are based on 10 000 replications.

3.4.2 Power simulations

3.4.2.1 A homogeneous nonlinear panel

We examine the empirical power under a modified PLSTAR(1) model because the transition function in (3.2) is replaced with $\tilde{F}(t;\gamma,c) = F(t;\gamma,c) + 0.5$. It is clear that $\tilde{F}(t) : \mathbb{R}_+ \to [0,1]$. Furthermore, the error term is assumed to be the same as in (3.18). The parameters in the modified PLSTAR model are assigned the following values

$$\pi_{i,10} = 0 \quad \forall i, \qquad \pi_{11} = 0.4, \qquad \pi_{i,20} = 1 \quad \forall i,$$

$$\pi_{21} \in \{0.4, 0.5, 0.55\}, \quad \gamma \in \{0.01, 1, 100\}, \quad c = T/2.$$

which generates a completely homogeneous panel. We first examine the power properties under an almost linear PLSTAR(1) model with $\gamma = 0.01$. Second, the power when the speed of transition in the PLSTAR(1) model may be characterized as intermediate with $\gamma = 1.00$ is investigated. Finally, we study the power when the transition takes place almost instantaneously with $\gamma = 100$, so the model practically contains a single structural break. Furthermore, within these three experiments the change in the intercept is set modest and equals 1, and the stationary root increases from 0.4 to 0.95 (assuming that a complete transition takes place). The design of this experiment implies panels with clear shift in levels and dynamics as long as $\gamma \geq 1$, see Figure 3.4 The results are presented in Tables 3.2-3.4.

When the DGP is an almost linear PLSTAR(1) model, we see from Table 3.2 that HT2 has the highest power, which is due to the near-linearity of the process. Test statistic HT2 is designed to have high power against linear models, and for small n and T it actually exhibits substantially higher power than in the case of a completely linear panel where the autoregressive coefficients range from 0.8 to 0.95, cf. Table 2b in Harris and Tzavalis (1999). This is natural because in our case the autoregressive parameter slowly grows e.g. from 0.4 to 0.67 over time (the case $T = 25$, $\gamma = 0.01$, and $\pi_{21} = 0.5$), rather than taking on e.g. the value 0.95 throughout the whole sample in a linear panel.[3] Statistic HT3 also performs better than our test. The reason is that HT3 is based on a more parsimonious alternative than our test which is penalized when the actual DGP is relatively simple.

For $T \geq 25$ and $n \geq 25$ our test performs satisfactorily and for $T \geq 50$ and all n (not reported here) all tests have unit power. In fact, increasing π_{21}

[3]For small T and $\gamma = 0.01$, a full transition from zero to one does not take place, see Figure 3.3. For example, $F(t = T = 25; \gamma = 0.01, c = 12.5) \approx 0.53$, implying that the value of the autoregressive parameter at the end of the period equals $\pi_{11} + 0.53\pi_{21} \in [0.61, 70]$.

Table 3.2 Empirical power of the test statistic in Corollary 3.3 and the tests in Harris and Tzavalis (1999). The DGP is a PAR(1) model.

$\gamma = 0.01$										
$\pi_{12} = 0.40$		$\pi_{21} = 0.40$			$\pi_{21} = 0.50$			$\pi_{21} = 0.55$		
T	n	T_1	HT2	HT3	T_1	HT2	HT3	T_1	HT2	HT3
5	5	0.05	0.35	0.08	0.05	0.27	0.07	0.04	0.23	0.06
5	10	0.06	0.54	0.10	0.05	0.41	0.09	0.05	0.34	0.08
5	25	0.06	0.86	0.16	0.05	0.72	0.12	0.05	0.63	0.11
5	50	0.06	0.99	0.23	0.06	0.93	0.17	0.05	0.87	0.15
5	100	0.07	1.00	0.38	0.07	1.00	0.24	0.06	0.99	0.21
10	5	0.14	0.80	0.28	0.12	0.69	0.21	0.10	0.62	0.19
10	10	0.18	0.97	0.41	0.14	0.91	0.30	0.11	0.85	0.26
10	25	0.28	1.00	0.72	0.20	1.00	0.56	0.16	0.99	0.47
10	50	0.41	1.00	0.93	0.27	1.00	0.81	0.21	1.00	0.70
10	100	0.61	1.00	1.00	0.39	1.00	0.97	0.30	1.00	0.93
25	5	0.68	1.00	0.97	0.58	1.00	0.89	0.52	1.00	0.83
25	10	0.89	1.00	1.00	0.79	1.00	0.99	0.74	1.00	0.98
25	25	1.00	1.00	1.00	0.98	1.00	1.00	0.96	1.00	1.00
25	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: The nominal size is 5%, and the results are based on 10 000 replications.

has only a moderate impact on power because the transition is very slow, see Figure 3.3.

Consider next the PLSTAR(1) model with $\gamma = 1$. In Table 3.3 we see that T_1 outperforms the other tests. It now becomes evident that our less parsimonious model is justified and is in fact necessary if one wants to capture the nonlinear behavior characterized by the PLSTAR(1) model. Our test actually has substantial power when the time dimension is as small as $T = 5, 10$, for almost all n and π_{21}. This is in contrast to HT2 and HT3, and especially the power of HT2 is very low.

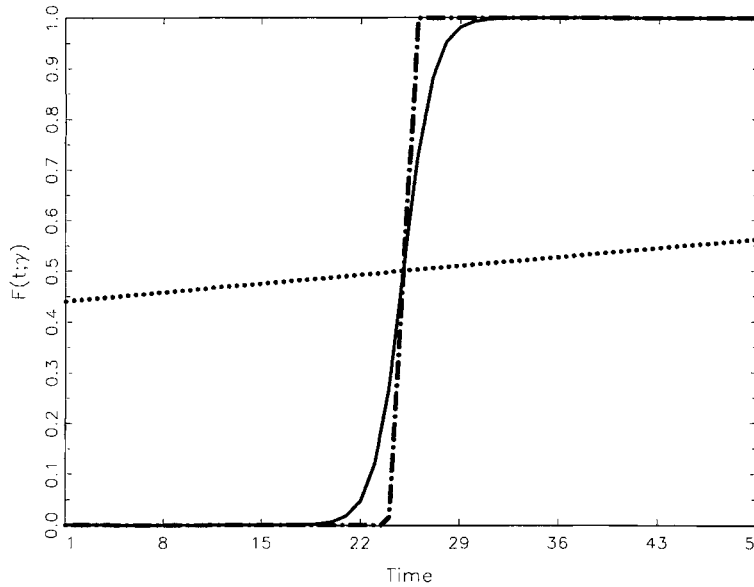
For $T \geq 25$ there are two points worth stressing. The first one is the successive break-down in power of the tests based on linear models by increasing π_{21}. To take an example, in the case $T = n = 50$ and $\pi_{21} = 0.40$ all tests have

Table 3.3 Empirical power of the test statistic in Corollary 3.3 and the tests in Harris and Tzavalis (1999). The DGP is a PLSTAR(1) model.

$\gamma = 1.00$										
$\pi_{11} = 0.40$		$\pi_{21} = 0.40$			$\pi_{21} = 0.50$			$\pi_{21} = 0.55$		
T	n	T_1	HT2	HT3	T_1	HT2	HT3	T_1	HT2	HT3
5	5	0.20	0.06	0.08	0.23	0.03	0.06	0.26	0.02	0.05
5	10	0.27	0.04	0.10	0.32	0.01	0.08	0.37	0.01	0.07
5	25	0.46	0.01	0.16	0.58	0.00	0.10	0.65	0.00	0.08
5	50	0.69	0.00	0.24	0.83	0.00	0.14	0.89	0.00	0.10
5	100	0.90	0.00	0.40	0.98	0.00	0.21	0.99	0.00	0.14
10	5	0.47	0.00	0.17	0.55	0.00	0.08	0.58	0.00	0.05
10	10	0.68	0.00	0.23	0.77	0.00	0.09	0.81	0.00	0.03
10	25	0.94	0.00	0.38	0.98	0.00	0.09	0.99	0.00	0.01
10	50	1.00	0.00	0.59	1.00	0.00	0.10	1.00	0.00	0.00
10	100	1.00	0.00	0.85	1.00	0.00	0.13	1.00	0.00	0.00
25	5	**0.79**	0.01	0.52	**0.76**	0.00	0.09	**0.74**	0.00	0.00
25	10	0.95	0.00	0.77	0.93	0.00	0.09	0.92	0.00	0.00
25	25	1.00	0.00	0.98	1.00	0.00	0.10	0.99	0.00	0.00
25	50	1.00	0.00	1.00	1.00	0.00	0.11	1.00	0.00	0.00
25	100	1.00	0.00	1.00	1.00	0.00	0.14	1.00	0.00	0.00
50	5	**0.97**	0.21	0.96	**0.77**	0.00	0.17	**0.57**	0.00	0.00
50	10	1.00	0.46	1.00	0.94	0.00	0.30	0.77	0.00	0.00
50	25	1.00	0.92	1.00	1.00	0.00	0.67	0.97	0.00	0.00
50	50	1.00	1.00	1.00	1.00	0.00	0.95	1.00	0.00	0.00
50	100	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00
100	5	**1.00**	1.00	1.00	**0.79**	0.00	0.76	**0.26**	0.00	0.00
100	10	1.00	1.00	1.00	1.00	0.00	0.97	0.35	0.00	0.00
100	25	1.00	1.00	1.00	1.00	0.00	1.00	0.56	0.00	0.00
100	50	1.00	1.00	1.00	1.00	0.00	1.00	0.79	0.00	0.00
100	100	1.00	1.00	1.00	1.00	0.00	1.00	0.96	0.00	0.00

Note: The nominal size is 5%, and the results are based on 10 000 replications.

Figure 3.3 The logistic transition function with different values of the speed of transition parameter: $\gamma = 0.01$ (dotted line), $\gamma = 1.00$ (solid line), and $\gamma = 100$ (dash-dotted line).

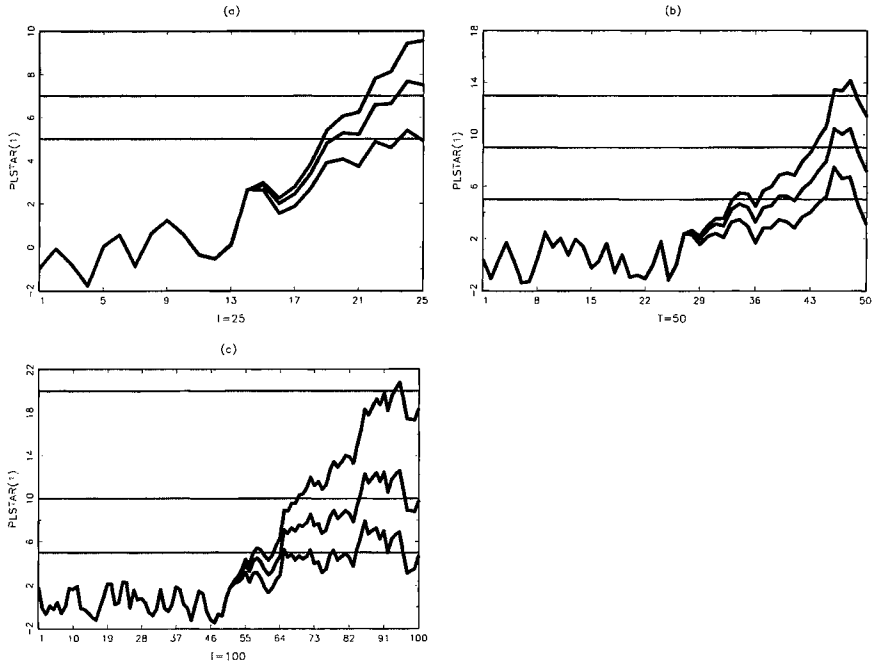


a power of unity. When π_{21} is increased to 0.50, a remarkable drop in power occurs for HT2 (from 1 to 0) whereas the power for T_1 and HT3 basically remains unchanged. Increasing π_{21} further to 0.55, there is a similar drop in power for HT3 (from 0.95 to 0) whereas our test still has power of unity. This emphasizes the relevance of the inclusion of the set of extra explanatory variables $\{t, ty_{i,t-1}\}$, missing from HT2, and $\{ty_{i,t-1}\}$ missing from HT3, in our auxiliary regression equation (3.3). The break-down in power could be explained by investigating the shape of the trajectories from the LSTAR(1) model as time evolves. In the cases $\pi_{21} = 0.40$, 0.50, and 0.55, the autoregressive parameter changes smoothly over time from 0.4 to 0.80, 0.90, and 0.95 , and typical realizations of the LSTAR model start at zero and end up at levels around 5, 9, and 13 respectively, see the panel (b) in Figure 3.4.[4] Thus, increasing π_{21} does not only imply that the LSTAR model is closer to being non-stable at the end of the period, it also implies a strive towards higher

[4]With $T = 50$ and $\gamma = 1$, a complete transition takes place, see Figure 3.3.

levels, see Chapter 1 for the discussion about the level leverage effect. These two effects contribute to the successive break-down in power of the HT2 and HT3 tests. The bias towards nonrejection for HT2 and HT3 tests when the time series in the panel have a pronounced shift in levels is demonstrated.

Figure 3.4 Typical realizations for a cross section unit in the PLSTAR(1) model where the sample sizes and the autoregressive roots in the nonlinear parts are varied.



The second phenomenon is that the power of T_1 actually decreases in T when $\pi_{21} = 0.55$ and $T \geq 25$, although it still increases monotonically with n. To study this, consider the case $n = 5$ while varying T and π_{21}. For $T = 25$ and $\pi_{21} = 0.40$, 0.50, and 0.55, the T_1 test show about the same power (≈ 0.77). In panel (a) in Figure 3.4, we can see that the trajectories at the end of the period reaches levels approximately equal to 5, 7, and 9 respectively. The differences in levels at the end of the period are modest. For $T = 50$ there are two obvious tendencies. First there is an evident drop in power for T_1 from 0.97 to 0.77 to 0.57 when π_{21} ranges from 0.40 to 0.55. The reason for this is that the levels of the trajectories at the end of the period now equal about 5,

9 and 13 respectively, see panel (b) in Figure 3.4, and more clear differences in the levels are encountered. Second, and perhaps more interesting, is that the power when $\pi_{21} = 0.55$ and $T = 50$ is lower than that when $\pi_{21} = 0.55$ and $T = 25$, i.e. 0.57 compared to 0.74. This drop in power is explained by a larger jump in level for $T = 50$ than for $T = 25$ (13 compared to 9) and that increasing the length of the time-series reveals the complexity of the process (cf. the (a) and (b) panels in Figure 3.4). It becomes clear that the term $ty_{t-1}$is not able to capture distinct changes in level and the autoregressive coefficient at the same time[5]. The same two tendencies, even more pronounced, can be observed when T is increased further. Specifically, when $T = 100$, the power decreases from 1.00 to 0.86 to 0.26 by increasing π_{21}, and in panel (c) in Figure 3.4, we see that the trajectories have now reached their long-run equilibriums, i.e. the levels at 5, 10, and 20 respectively.[6] Furthermore, the reduction in power from 0.57 (the case $T = 50$ and $\pi_{21} = 0.55$) to 0.26 (the case $T = 100$ and $\pi_{21} = 0.55$) is larger than the reduction in power from 0.74 (the case $T = 25$ and $\pi_{21} = 0.55$) to 0.57 (the case $T = 50$ and $\pi_{21} = 0.55$). Larger reductions in power by increasing T can be explained by the fact that the levels at the end of the sample periods for $T = 25$, 50, and 100, equal 9, 13, and 20 respectively.

Moreover, it should also be mentioned that increasing T further results in yet another reduction in power. This reduction in power continues actually until $T = 250$ (not reported here), from where the power rapidly increases and reaches unity. This can then be seen as a measure of that we need T (sufficiently) larger than 250 for the term ty_{t-1} to adequately capture the nonlinear structure of an LSTAR(1) model allowing for a modest shift in the intercept and an almost non-stable root at the end of the sample period. We conclude that despite a panel data model with an increased information set, there may still be a need for a third-order approximation of the DGP (3.1) and (3.2) to achieve acceptable power when the cross section dimension of the panel is small.

From Table 3.4 we can see that when the DGP is a PLSTAR(1) model behaving almost like a PTAR(1) model, the empirical powers of the tests are lower than in the previous case. Differences in power compared to what is reported in Table 3.3 are modest, however, and the response of the tests is robust against the change in γ from 1 to 100 in the PLSTAR model.

[5]Despite a panel set-up and the large amount of information available, these results indicate that a test with a a third-order Taylor approximation of the transition function in (3.2) might be preferable. However, this will render the analytical tractability of the results derived in Section 3.3.

[6]These long-run equilibriums are given by $(\pi_{10} + \pi_{20})/(1 - \pi_{11} - \pi_{21})$.

Table 3.4 Empirical power of the test statistic in Corollary 3.3 and the tests in Harris and Tzavalis (1999). The DGP is a PTAR(1) model.

$\gamma = 100$										
$\pi_{11} = 0.40$		$\pi_{21} = 0.40$			$\pi_{21} = 0.50$			$\pi_{21} = 0.55$		
T	n	T_1	HT2	HT3	T_1	HT2	HT3	T_1	HT2	HT3
5	5	0.21	0.01	0.04	0.25	0.01	0.02	0.28	0.00	0.02
5	10	0.28	0.00	0.04	0.35	0.01	0.02	0.39	0.01	0.01
5	25	0.45	0.00	0.03	0.58	0.00	0.01	0.65	0.00	0.00
5	50	0.65	0.00	0.02	0.81	0.00	0.01	0.86	0.00	0.00
5	100	0.87	0.00	0.01	0.97	0.00	0.00	0.98	0.00	0.00
10	5	0.36	0.00	0.10	0.42	0.00	0.04	0.44	0.00	0.01
10	10	0.51	0.00	0.12	0.59	0.00	0.03	0.65	0.00	0.03
10	25	0.81	0.00	0.15	0.89	0.00	0.01	0.93	0.00	0.01
10	50	0.97	0.00	0.21	0.99	0.00	0.00	1.00	0.00	0.00
10	100	1.00	0.00	0.34	1.00	0.00	0.00	1.00	0.00	0.00
25	5	0.67	0.00	0.41	0.59	0.00	0.04	0.57	0.00	0.00
25	10	0.86	0.00	0.65	0.80	0.00	0.03	0.76	0.00	0.00
25	25	1.00	0.00	0.95	0.98	0.00	0.02	0.97	0.00	0.00
25	50	1.00	0.00	1.00	1.00	0.00	0.02	1.00	0.00	0.00
25	100	1.00	0.00	1.00	1.00	0.00	0.02	1.00	0.00	0.00
50	5	0.96	0.20	0.95	0.65	0.00	0.13	0.44	0.00	0.00
50	10	1.00	0.42	1.00	0.86	0.00	0.24	0.62	0.00	0.00
50	25	1.00	0.91	1.00	0.99	0.00	0.56	0.90	0.00	0.00
50	50	1.00	1.00	1.00	1.00	0.00	0.89	0.99	0.00	0.00
50	100	1.00	1.00	1.00	1.00	0.00	0.99	1.00	0.00	0.00
100	5	1.00	0.00	1.00	0.81	0.00	0.74	0.21	0.00	0.00
100	10	1.00	0.00	1.00	0.97	0.00	0.92	0.26	0.00	0.00
100	25	1.00	0.00	1.00	1.00	0.00	1.00	0.41	0.00	0.00
100	50	1.00	0.00	1.00	1.00	0.00	1.00	0.63	0.00	0.00
100	100	1.00	0.00	1.00	1.00	0.00	1.00	0.85	0.00	0.00

Note: The nominal size is 5%, and the results are based on 10 000 replications.

3.4.2.2 A heterogeneous nonlinear panel

A less restricted approach is adopted in the next two experiments because a heterogeneous panel is considered. In the first of these experiments, this is achieved by specifying the following parameter values in the PLSTAR model

$$\begin{aligned} &\pi_{i,11} = 0 \;\; \forall i, \quad \pi_{12} = 0.4, \quad \pi_{i,20} \sim U\left[0.5, 1.5\right], \\ &\pi_{21} = 0.5, \qquad c = T/2, \qquad \gamma = 1. \end{aligned} \tag{3.19}$$

The cross section specific parameter $\pi_{i,20}$ is drawn once from the uniform distribution and thereafter held fixed throughout the replications. By doing this we allow for cross section specific long-run attractors.[7]

In the second experiment we increase the heterogeneity of our PLSTAR model allowing π_{21} to be individual-specific as well, denoted $\pi_{i,21}$. We choose the same design as in (3.19) but let $\pi_{i,21} \sim U\left[0.5, 0.55\right]$.[8] The empirical powers for these two experiments are reported in Table 3.5.

In Table 3.5 we can see that the power for T_1 is satisfactory for all combinations of T and n, whereas HT3 requires T, $n \geq 50$ or $T = 100$ and $n \geq 5$ to achieve a substantial power. In addition, the power for HT3 (the power for HT2 remains zero) is clearly reduced compared to the results in Table 3.3 with $\pi_{21} = 0.5$. For instance, when $T = n = 50$ the power is reduced by 15% (from 0.95 to 0.81) in the first experiment and by 89% (from 0.95 to 0.10) in the second experiment where the heterogeneity is increased, whereas the power of our test is still unity. In general the reduction is much smaller for our test. This is of course an important aspect for practitioners, and we see that the T_1 test appears more robust than the HT3 test when each individual is allowed for its own long-run attractor, even though the design of the DGP's is not supported by the model specification in (3.1) with (3.2).

3.4.2.3 A heterogeneous linear panel

In the next experiment we examine the power when the DGP is either a stationary (S) or trend stationary (TS) process defined according to the DGP's under the alternative in Harris and Tzavalis (1999). Thus,

$$y_{it} = \alpha_i + \varphi y_{i,t-1} + u_{it}, \quad \text{(S)},$$

[7]The long-run attractor for individual i is given by $(\pi_{10} + \pi_{i,20})/(1 - \pi_{11} - \pi_{21}) \in [5, 15]$.

[8]The long-run attractor for individual i is now given by $(\pi_{10} + \pi_{i,20})/(1 - \pi_{11} - \pi_{i,21}) \in [5, 30]$. Furthermore, notice that the model specification in (3.1) with (3.2) only supports heterogenity in $\pi_{i,10}$ and $\pi_{i,20}$.

Table 3.5 Empirical power of the test statistic in Corollary 3.3 and the tests in Harris and Tzavalis (1999). Heterogeneous panels.

T	n	T_1	HT2	HT3
25	5	0.73 (0.73)	0.00 (0.00)	0.07 (0.04)
25	10	0.92 (0.91)	0.00 (0.00)	0.07 (0.02)
25	25	1.00 (1.00)	0.00 (0.00)	0.06 (0.00)
25	50	1.00 (1.00)	0.00 (0.00)	0.04 (0.00)
25	100	1.00 (1.00)	0.00 (0.00)	0.02 (0.00)
50	5	0.71 (0.62)	0.00 (0.00)	0.17 (0.05)
50	10	0.87 (0.84)	0.00 (0.00)	0.25 (0.06)
50	25	1.00 (0.98)	0.00 (0.00)	0.49 (0.08)
50	50	1.00 (1.00)	0.00 (0.00)	0.81 (0.10)
50	100	1.00 (1.00)	0.00 (0.00)	0.98 (0.11)
100	5	0.77 (0.64)	0.00 (0.00)	0.65 (0.34)
100	10	0.93 (0.75)	0.00 (0.00)	0.89 (0.39)
100	25	1.00 (0.97)	0.00 (0.00)	1.00 (0.74)
100	50	1.00 (1.00)	0.00 (0.00)	1.00 (0.93)
100	100	1.00 (1.00)	0.00 (0.00)	1.00 (1.00)

Notes: The nominal size is 5%, and the results are based on 10 000 replications. Values in parentheses correspond to the power from a panel with a higher degree of heterogeneity.

or

$$y_{it} = \alpha_i + \alpha_i(1-\varphi)t + \varphi y_{i,t-1} + u_{it}, \quad \text{(T.S.)},$$

where α_i is drawn once from the standard normal distribution and $\varphi \in \{0.80, 0.90, 0.95\}$ and thereafter held fixed throughout the replications. The null hypothesis is the DGP in (3.18).

The results in Table 3.6 show, as may be expected, that T_1 is clearly inferior to HT2 for all n and T. However, our test is reasonably powerful for $\varphi = 0.80$, 0.90, and $T = 50$ and $n \geq 25$, or $T = 100$ and all n. However, for $\varphi = 0.95$ we see that our test has power close to the nominal size. The situation is different when we instead consider a trend stationary DGP's and compare the power of T_1 to HT3. We see that for $n = 5$ and $T \geq 5$ the differences in power between T_1 and HT3 are marginal . In fact, studying the same situation but assuming that $n > 5$ and $T \geq 5$, our test T_1 actually has higher power than the HT3 test, and in particular, for $\varphi = 0.95$, T_1 performs substantially better

Table 3.6 Empirical power of the test statistic in Corollary 3.3 and the tests in Harris and Tzavalis (1999). The DGP's are stationary and trend stationary.

		$\varphi = 0.80$				$\varphi = 0.90$				$\varphi = 0.95$			
		S		T.S.		S		T.S.		S		T.S.	
		T_1	HT2	T_1	HT3	T_1	HT2	T_1	HT3	T_1	HT2	T_1	HT3
T	n												
5	5	0.05	0.29	0.08	0.06	0.05	0.18	0.08	0.05	0.05	0.13	0.09	0.05
5	10	0.03	0.41	0.13	0.07	0.04	0.22	0.11	0.05	0.04	0.14	0.14	0.05
5	25	0.03	0.69	0.42	0.08	0.03	0.39	0.32	0.06	0.04	0.17	0.41	0.05
5	50	0.01	0.91	0.77	0.09	0.02	0.51	0.61	0.06	0.03	0.23	0.70	0.06
5	100	0.00	1.00	0.98	0.11	0.01	0.76	0.95	0.06	0.02	0.35	0.93	0.06
10	5	0.09	0.50	0.11	0.11	0.06	0.30	0.07	0.08	0.06	0.19	0.07	0.07
10	10	0.08	0.70	0.17	0.14	0.05	0.40	0.13	0.08	0.05	0.23	0.11	0.07
10	25	0.08	0.94	0.44	0.20	0.04	0.64	0.29	0.09	0.04	0.34	0.27	0.07
10	50	0.08	0.99	0.82	0.30	0.03	0.87	0.62	0.10	0.02	0.48	0.52	0.07
10	100	0.08	1.00	0.99	0.47	0.02	0.98	0.94	0.13	0.02	0.71	0.89	0.07
25	5	0.29	0.93	0.34	0.43	0.13	0.59	0.15	0.17	0.09	0.36	0.11	0.11
25	10	0.37	0.99	0.62	0.61	0.14	0.79	0.24	0.21	0.08	0.47	0.13	0.11
25	25	0.59	1.00	0.98	0.92	0.17	0.97	0.55	0.31	0.06	0.73	0.31	0.11
25	50	0.82	1.00	1.00	0.99	0.19	1.00	0.91	0.48	0.06	1.00	0.63	0.14
25	100	0.97	1.00	1.00	1.00	0.25	1.00	1.00	0.72	0.04	1.00	0.94	0.19
50	5	0.86	1.00	0.91	0.95	0.29	0.92	0.36	0.43	0.14	0.58	0.16	0.18
50	10	1.00	1.00	1.00	0.99	0.38	0.99	0.64	0.63	0.15	0.78	0.25	0.20
50	25	1.00	1.00	1.00	1.00	0.60	1.00	0.98	0.91	0.18	0.97	0.58	0.32
50	50	1.00	1.00	1.00	1.00	0.81	1.00	1.00	0.99	0.20	1.00	0.90	0.46
50	100	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.27	1.00	1.00	0.72
100	5	0.98	1.00	1.00	1.00	0.67	1.00	0.91	0.94	0.28	0.93	0.38	0.42
100	10	1.00	1.00	1.00	1.00	0.87	1.00	1.00	1.00	0.38	0.99	0.65	0.62
100	25	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.59	1.00	0.98	0.91
100	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.81	1.00	1.00	1.00
100	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00

Note: The nominal size is 5%, and the results are based on 10 000 replications.

than the latter test statistic. We may conclude that our test seems to have reasonable power properties against stationary/trend stationary alternatives, and that the power of T_1 approaches unity at a faster rate with n than the power of HT3.

3.4.2.4 Power when viewing T and n as asymptotic when the true T is finite

The last experiment concerns the empirical power of the test in Corollary 3.4 (denoted T_2) when we treat both n and T as asymptotic but T is actually finite. This also demonstrates the mean-shift effect and variance effect addressed in Section 3.3. For comparison the same effects are investigated for the corresponding tests by Harris and Tzavalis (1999), denoted $\mathrm{HT2}^*$ and $\mathrm{HT3}^*$. For this study we choose the same set-up as for the experiment with a homogeneous panel and $\gamma = 1.00$. These findings are presented in Table 3.7.

In this table we see that treating both T and n as asymptotic leads to rather severe size distortions and the statistic T_2 becomes undersized for almost all n and $T \leq 25$. This indicates that the variance effect discussed in Section 3.3 dominates the mean shift effect, resulting in a net reduction in the seize of the test, which agree with the findings reported in Harris and Tzavalis (1999). As a result we see a clear drop in power (cf. Table 3.3). For $T \geq 50$ and all n the power reduction is marginal. It seems that in order to maintain a test with correct size when letting n and T tend to infinity, a diagonal convergence criterion should be imposed to control for the mean-shift and variance effect.

3.5 Conclusions

In this chapter we argue that many of the traditional panel data unit root tests are based on too restrictive panels because it is likely that e.g. macroeconomic panels involve cross sections with time series that exhibit structural changes in levels. We also emphasize the importance of testing unit roots in a nonlinear panel which accommodates a smooth shift in levels and the dynamic structure (the PLSTAR model) because here the conventional unit root tests, such as the test in Harris and Tzavalis (1999), are biased towards nonrejection of the null hypothesis.

The unit root test that we derive in the PLSTAR model is based on an auxiliary regression, and inference is based on the LSDV estimator under the assumption that the disturbances are independently and identically distributed under the null hypothesis. It is shown that the test statistic is normally distributed where the first two moments are calculated analytically. Due to

Table 3.7 Empirical power and size of the test statistic in Corollary 3.4 and the tests in Harris and Tzavalis (1999), assuming that T and n are large.

$\gamma = 1.00$									
$\pi_{11} = 0.40$			$\pi_{21} = 0.40$			$\pi_{21} = 0.50$			$\pi_{21} = 0.55$
		Size	Power			Power			Power
T	n	T_2	T_2	HT2*	HT3*	T_2	HT2*	HT3*	T_2
5	5	0.00	0.02	0.01	0.00	0.02	0.00	0.00	0.02
5	10	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01
5	25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
5	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
5	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
10	5	0.01	0.21	0.00	0.01	0.26	0.00	0.00	0.27
10	10	0.01	0.34	0.00	0.00	0.76	0.00	0.00	0.47
10	25	0.00	0.63	0.00	0.00	0.96	0.00	0.00	0.83
10	50	0.00	0.88	0.00	0.00	1.00	0.00	0.00	0.99
10	100	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00
25	5	0.05	0.72	0.01	0.32	0.66	0.00	0.02	0.64
25	10	0.04	0.92	0.00	0.51	0.88	0.00	0.02	0.86
25	25	0.02	1.00	0.00	0.89	1.00	0.00	0.01	0.99
25	50	0.02	1.00	0.00	1.00	1.00	0.00	0.00	1.00
25	100	0.00	1.00	0.0	01.00	1.00	0.00	0.00	1.00
50	5	0.08	0.96	0.18	0.95	0.73	0.00	0.11	0.53
50	10	0.06	1.00	0.38	1.00	0.91	0.00	0.18	0.72
50	25	0.04	1.00	0.87	1.00	1.00	0.00	0.43	0.95
50	50	0.03	1.00	1.00	1.00	1.00	0.00	0.79	1.00
50	100	0.02	1.00	1.00	1.00	1.00	0.00	0.99	1.00
100	5	0.08	1.00	1.00	1.00	0.85	0.00	0.72	0.23
100	10	0.07	1.00	1.00	1.00	0.97	0.00	0.95	0.31
100	25	0.05	1.00	1.00	1.00	1.00	0.00	0.00	0.51
100	50	0.04	1.00	1.00	1.00	1.00	0.00	0.00	0.74
100	100	0.03	1.00	1.00	1.00	1.00	0.00	0.00	0.94

Note: The nominal size is 5%, and the results are based on 10 000 replications.

the fact that these moments are known for a fixed sample size, we are able to analyze how this will affect the inference when T is infinite.

Finite-sample properties of the test are explored through Monte Carlo simulations and show satisfactory results. The size distortion is modest and the power is generally superior to the power of the tests in Harris and Tzavalis (1999). We especially demonstrate that the traditional tests in Harris and Tzavalis (1999) lack power if the shift in levels is too evident, but with more modest shifts the test with a linear trend has reasonable power. There are situations, however, in which our test may be modified in order to increase the power. This can be done by applying a higher-order Taylor approximation to the PLSTAR model, but it will make the analytical results less tractable, and the question is left for further research.

Appendix A

Proofs

Lemma 1 *Let* $\mathbf{M}_T = [m_{ij}]_{T\times T}$ *be the matrix defined in (3.5). Then the* (i,j)*-th element of* $\mathbf{M}_T$ *can be expressed*

$$m_{ij} = \frac{2}{T(T^2-1)}\left((T+1)\left((2T+1)-3i-3j\right)+6ij\right), \tag{A.1}$$

for any $i,j = 1,...,T$.

Proof. Note that the inverse $(\mathbf{X}_T'\mathbf{X}_T)^{-1} = \frac{2}{T(T-1)}\begin{bmatrix} (2T+1) & -3 \\ -3 & 6(T+1)^{-1} \end{bmatrix}$. Thus, the formula for m_{ij} in (A.1) holds by computing $\mathbf{M}_T$. ■

Lemma 2 *Let* $\mathbf{C}_T$ *be the* $(T\times T)$ *matrix defined in (3.7), and* $\mathbf{D}_T$ *and* $\mathbf{M}_T$ *be the* $(T\times T)$ *matrices defined in (3.5). Then*

$$\begin{aligned} &\textit{the (i,j)-th element of } \mathbf{C}_T'\mathbf{M}_T\mathbf{C}_T \\ &= \begin{cases} \sum_{t=i+1}^{T}\sum_{s=j+1}^{T} m_{ts}, & i,j = 1,...,T-1. \\ 0, & i=T \ \ or \ \ j=T. \end{cases}, \end{aligned} \tag{A.2}$$

$$\begin{aligned} &\textit{the (i,j)-th element of } \mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T \\ &= \begin{cases} \sum_{t=i+1}^{T}\sum_{s=j+1}^{T} tsm_{ts}, & i,j = 1,...,T-1. \\ 0, & i=T \ \ or \ \ j=T. \end{cases}, \end{aligned} \tag{A.3}$$

$$\begin{aligned} &\textit{the (i,j)-th element of } \mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T \\ &= \begin{cases} \sum_{t=i+1}^{T}\sum_{s=j+1}^{T} sm_{ts}, & i,j = 1,...,T-1. \\ 0, & i=T \ \ or \ \ j=T. \end{cases}, \end{aligned} \tag{A.4}$$

$$\begin{aligned} & \textit{the (i,j)-th element of } \mathbf{C}_T'\mathbf{M}_T \\ = & \begin{cases} \sum_{t=i+1}^{T} m_{tj}, & , i=1,...,T-1,\ j=1,...,T. \\ 0, & i=T. \end{cases} , \end{aligned} \tag{A.5}$$

$$\begin{aligned} & \textit{the (i,j)-th element of } \mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T \\ = & \begin{cases} \sum_{t=i+1}^{T} t m_{tj}, & , i=1,...,T-1, j=1,...,T. \\ 0, & i=T. \end{cases} . \end{aligned} \tag{A.6}$$

Proof. Applying (A.1) to $\mathbf{C}_T'\mathbf{M}_T\mathbf{C}_T, \mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T, \mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T$, $\mathbf{C}_T'\mathbf{M}_T$, and $\mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T$, respectively, gives the formulas in (A.2)-(A.6). ■

Lemma 3 . *Consider model (3.3) when (3.4) and (A1)-(A3) in Assumption 1 hold. Furthermore, let $W_{i,jt}$, $j=1,...,5$, be defined in (3.5). Then, for any fixed T,*

$$\underset{n\to\infty}{p\lim} n^{-1}\sum_{i=1}^{n} W_{i,1t} = \frac{1}{15}(T^2-4)\sigma_u^2, \tag{A.7}$$

$$\underset{n\to\infty}{p\lim} n^{-1}\sum_{i=1}^{n} W_{i,2t} = \frac{1}{420}(T^2-4)(11T^2+14T-1)\sigma_u^2, \tag{A.8}$$

$$\underset{n\to\infty}{p\lim} n^{-1}\sum_{i=1}^{n} W_{i,3t} = \frac{1}{30}(T^2-4)(T+1)\sigma_u^2, \tag{A.9}$$

$$\underset{n\to\infty}{p\lim} n^{-1}\sum_{i=1}^{n} W_{i,4t} = -\frac{1}{2}(T-2)\sigma_u^2, \tag{A.10}$$

$$\underset{n\to\infty}{p\lim} n^{-1}\sum_{i=1}^{n} W_{i,5t} = -\frac{1}{60}(17T+19)(T-2)\sigma_u^2. \tag{A.11}$$

Proof. (i) Under (3.4), express the model in (3.3) as

$$\mathbf{y}_{i,-1} = y_{i0}\boldsymbol{\iota}_T + \mathbf{C}_T\mathbf{u}_i. \tag{A.12}$$

(ii) Note that $\mathbf{Q}_T$ defined in (3.5) is orthogonal to $\boldsymbol{\iota}_T$ so pre-multiplying $\mathbf{Q}_T$ on both sides of $\mathbf{y}_{i,-1}$ in (A.12) yields

$$\mathbf{Q}_T\mathbf{y}_{i,-1} = \mathbf{Q}_T\mathbf{C}_T\mathbf{u}_i. \tag{A.13}$$

Since $\mathbf{Q}_T$ is idempotent it follows from (A.13) that

$$\begin{aligned} W_{i,1t} &= \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{y}_{i,-1} \\ &= \mathbf{u}_i'\mathbf{C}_T'\mathbf{C}_T\mathbf{u}_i - \mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{C}_T\mathbf{u}_i. \end{aligned} \tag{A.14}$$

In (A.14), we have

$$\mathbf{u}_i'\mathbf{C}_T'\mathbf{C}_T\mathbf{u}_i = \sum_{t=1}^{T}(T-t)u_{it}^2 + 2\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(T-s)u_{it}u_{is},$$

and

$$\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{C}_T\mathbf{u}_i = \sum_{t=2}^{T}(\sum_{i=t}^{T}\sum_{j=t}^{T}m_{ij})u_{i,t-1}^2 + 2\sum_{t=2}^{T}\sum_{s=t+1}^{T}(\sum_{i=t}^{T}u_{i,t-1}\sum_{j=s}^{T}m_{ij})u_{i,s-1}.$$

By assumption (A1), $\{u_{it}\}$ is a sequence of i.i.d. random variables across i and t so that

$$\begin{aligned} E(W_{i,1t}) &= E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{C}_T\mathbf{u}_i) - E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{C}_T\mathbf{u}_i) \\ &= \sum_{t=1}^{T}(T-t)\sigma_u^2 - \sum_{t=2}^{T}(\sum_{i=t}^{T}\sum_{j=t}^{T}m_{ij})\sigma_u^2, \end{aligned}$$

and it follows now from (A.2) that $E(W_{i,1t}) = \frac{1}{15}(T^2-4)\sigma_u^2$. Finally, by the LLN, (A.7) holds. Similarly, we can show that (A.8)-(A.11) hold. ■

Proof of Theorem 3.1. (i) It follows from Anderson and Hsiao (1981) that the LSDV estimators of ρ and ϕ for model (3.3), under the null hypothesis (3.4), equal

$$\begin{aligned} \begin{pmatrix} \widehat{\rho}-1 \\ \widehat{\phi} \end{pmatrix} &= \left[\sum_{i=1}^{n}\begin{pmatrix} \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{y}_{i,-1} & \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{D}_T\mathbf{y}_{i,-1} \\ \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{y}_{i,-1} & \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{D}_T\mathbf{y}_{i,-1} \end{pmatrix}\right]^{-1} \\ &\quad \times \left[\sum_{i=1}^{n}\begin{pmatrix} \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{u}_i \\ \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{u}_i \end{pmatrix}\right], \end{aligned} \tag{A.15}$$

where $\mathbf{Q}_T$ and $\mathbf{D}_T$ are defined in (3.5). It follows immediately from (A.15) that the expression for $(\widehat{\rho}-1)$ in (3.5) holds.

(ii) It follows from Lemma 3, (A.12), and (A.13) that the probability of the numerator of (A.15) is

$$\underset{n\to\infty}{p\lim} n^{-1}\sum_{i=1}^{n}\begin{pmatrix} \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{u}_i \\ \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{u}_i \end{pmatrix} = -\sigma_u^2\begin{pmatrix} \frac{1}{2}(T-2) \\ \frac{1}{60}(17T+19)(T-2) \end{pmatrix},$$

whereas the probability limit of denominator in (A.15) equals

$$\begin{aligned} &\underset{n\to\infty}{p\lim} n^{-1}\sum_{i=1}^{n}\begin{pmatrix} \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{y}_{i,-1} & \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{D}_T\mathbf{y}_{i,-1} \\ \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{y}_{i,-1} & \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{D}_T\mathbf{y}_{i,-1} \end{pmatrix} \\ &= \sigma_u^2\begin{pmatrix} \frac{1}{15}(T^2-4) & \frac{1}{30}(T^2-4)(T+1) \\ \frac{1}{30}(T^2-4)(T+1) & \frac{1}{420}(T^2-4)(11T^2+14T-1) \end{pmatrix}. \end{aligned}$$

(iii) By the Slutsky Theorem (see, e.g. page 286 of Davidson (1994)) we obtain

$$
\begin{aligned}
& \underset{n\to\infty}{p\lim} \begin{pmatrix} \widehat{\rho}-1 \\ \widehat{\phi} \end{pmatrix} \\
= & \left[\underset{n\to\infty}{p\lim} n^{-1} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{y}_{i,-1} & \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{D}_T\mathbf{y}_{i,-1} \\ \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{y}_{i,-1} & \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{D}_T\mathbf{y}_{i,-1} \end{pmatrix} \right]^{-1} \\
& \times \left[\underset{n\to\infty}{p\lim} n^{-1} \begin{pmatrix} \sum_{i=1}^{n} \mathbf{y}_{i,-1}'\mathbf{Q}_T\mathbf{u}_i \\ \mathbf{y}_{i,-1}'\mathbf{D}_T\mathbf{Q}_T\mathbf{u}_i \end{pmatrix} \right] \\
= & \begin{bmatrix} -0.25\left(23T^2-21T-74\right)/\left((T^2-2)(T+2)\right) \\ -3.5/\left(T^2-2\right) \end{bmatrix}. \qquad \text{(A.16)}
\end{aligned}
$$

provided that $T > 2$. The inconsistency of $(\widehat{\rho}-1)$ in (3.6) is the upper element of the limit vector on the right-hand side of (A.16). ■

Lemma 4 *Let* $\mathbf{M} = [m_{ij}]_{T\times T}$ *and* $\mathbf{N} = [n_{ij}]_{T\times T}$ *be any constant matrices. Assume that (A1) and (A3) in Assumption 1 hold. Then,*

$$
\begin{aligned}
& E(\mathbf{u}_i'\mathbf{M}\mathbf{u}_i)(\mathbf{u}_i'\mathbf{N}\mathbf{u}_i) \\
= & \ \mu_4 \sum_{t=1}^{T} m_{tt}n_{tt} + \sigma_u^4\{\sum_{t=1}^{T}[m_{tt}(\sum_{s=t+1}^{T} n_{ss})] + \sum_{t=1}^{T}[n_{tt}(\sum_{s=t+1}^{T} m_{ss})] \\
& + \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(m_{ts}n_{ts}) + \sum_{s=1}^{T-1}\sum_{t=s+1}^{T}(m_{ts}n_{ts}) \\
& + \sum_{t=1}^{T-1}[\sum_{s=t+1}^{T}(m_{ts}n_{st})] + \sum_{s=1}^{T-1}[\sum_{t=s+1}^{T}(m_{ts}n_{st})]\}.
\end{aligned}
$$

Proof. Because u_{it} is i.i.d., $\mathbf{u}_i'\mathbf{M}\mathbf{u}_i = \sum_{s=1}^{T}\sum_{t=1}^{T} m_{st}u_{is}u_{it}$, and $\mathbf{u}_i'\mathbf{M}\mathbf{u}_i = \sum_{s=1}^{T}\sum_{t=1}^{T} n_{st}u_{is}u_{it}$, it follows by collecting all the terms of $m_{tt}n_{tt}$ for $t = 1,...,T$ in $(\mathbf{u}_i'\mathbf{M}\mathbf{u}_i)(\mathbf{u}_i'\mathbf{N}\mathbf{u}_i)$ and now taking the expectations we obtain the coefficient of μ_4. For the coefficients of σ_u^4, considering all the terms of $m_{ts}n_{ts}$, $m_{ts}n_{st}$ and $m_{tt}n_{ss}$, for $t \neq s$, in $(\mathbf{u}_i'\mathbf{M}\mathbf{u}_i)(\mathbf{u}_i'\mathbf{N}\mathbf{u}_i)$ and computing all the expectations for those terms yield the coefficients for σ_u^4. ■

Lemma 5 *Consider model (3.3) when (3.4) and (A1)-(A3) in Assumption 1 hold. Let* $W_{i,jt}$, $j = 1,...,5$, *be defined in (3.5). Then, for any fixed* $T > 1$,

$$
\begin{aligned}
E(W_{i,1t})^2 = & \ \frac{1}{210}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\} \times \{\mu_4(T^4-25) \\
& + \sigma_u^4[\frac{1}{10}(13T^5-30T^4+10T^3-23T+750)]\}, \qquad \text{(A.17)}
\end{aligned}
$$

$$
\begin{aligned}
E(W_{i,2t})^2 &= \frac{1}{1081\,080}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\} \times \{\mu_4(1382T^8+3185T^7 \\
&\quad +4141T^6+3770T^5-80\,204T^4-153\,907T^3+163\,997T^2 \\
&\quad +343\,512T-2484)+\sigma_u^4[\frac{1}{210}(244\,231T^9-282\,072T^8 \\
&\quad -1261\,143T^7-1695\,918T^6-7526\,883T^5+40\,810\,812T^4 \\
&\quad +99\,837\,023T^3-95\,101\,902T^2-215\,126\,028T \\
&\quad +1564\,920)]\}, \qquad \text{(A.18)} \\
E(W_{i,3t})^2 &= \frac{1}{13\,860}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\} \times \{\mu_4(24T^6+33T^5+32T^4 \\
&\quad -926T^2-825T+1182)+\sigma_u^4[\frac{1}{30}(715T^7-873T^6-902T^5 \\
&\quad -1890T^4-7205T^3+81\,063T^2+78\,672T-106\,380)]\}, \qquad \text{(A.19)} \\
E(W_{i,4t})^2 &= \frac{1}{105}\left\{(T-2)\,/\,\left(T(T^2-1)\right)\right\} \times \{\mu_4\left(29T^3-47T^2-23T+59\right) \\
&\quad +\sigma_u^4[\frac{1}{2}\left(56T^4-167T^3+226T^2+131T-354\right)]\}, \qquad \text{(A.20)} \\
E(W_{i,5t})^2 &= \frac{1}{630}\left\{(T-2)\,/\,\left(T(T^2-1)\right)\right\} \times \{\mu_4(64T^5+29T^4-37T^3+16T^2 \\
&\quad -357T-435)+\sigma_u^4[\frac{1}{20}(1153T^6-904T^5-2840T^4-3130T^3 \\
&\quad -1013T^2+23\,834T+26\,100)]\}. \qquad \text{(A.21)}
\end{aligned}
$$

Proof. For instance, by the definition of $W_{i,3t}$ we can write

$$
\begin{aligned}
E(W_{i,3t})^2 &= E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)^2-2E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i) \\
&\quad +E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)^2. \qquad \text{(A.22)}
\end{aligned}
$$

From the assumptions on u_{it} and Lemmas 2 and 4 we are able to compute the first term of the right-hand side of (A.22),

$$
\begin{aligned}
& E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)^2 \\
= \; & \mu_4[\sum_{t=1}^{T}(\sum_{j=1}^{T}j-\sum_{i=1}^{t}i)^2]+\sigma_u^4\{2\sum_{t=1}^{T}[(\sum_{j=1}^{T}j-\sum_{i=1}^{t}i)(\sum_{s=t+1}^{T}(\sum_{j=1}^{T}j-\sum_{i=1}^{s}i))] \\
& +4\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}[(\sum_{j=1}^{T}j-\sum_{i=1}^{s}i)^2]\} \\
= \; & \mu_4[\frac{1}{60}T\left(T^2-1\right)\left(8T^2+5T-2\right)] \\
& +\sigma_u^4[\frac{1}{180}T\left(T^2-1\right)(T-2)\left(50T^2+22T-21\right)], \qquad \text{(A.23)}
\end{aligned}
$$

for the second term of the right-hand side of (A.22), we have

$$
\begin{aligned}
& E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i) \\
= \; & \mu_4[\sum_{t=1}^{T}(\sum_{j=1}^{T} j - \sum_{k=1}^{t} k)(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T} jm_{kj})] \\
& +\sigma_u^4\{\sum_{t=1}^{T}[(\sum_{j=1}^{T} j - \sum_{k=1}^{t} k)(\sum_{s=t+2}^{T}(\sum_{k=s}^{T}\sum_{j=s}^{T} jm_{kj}))] \\
& +\sum_{t=1}^{T}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T} jm_{kj})(\sum_{s=t+1}^{T}(\sum_{j=1}^{T} j - \sum_{k=1}^{s} k))] \\
& +2\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}[(\sum_{j=1}^{T} j - \sum_{k=1}^{s} k)(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T} jm_{kj})] \\
& +2\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}[(\sum_{j=1}^{T} j - \sum_{k=1}^{t} k)(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T} jm_{kj})]\} \\
= \; & \mu_4[\frac{1}{120}(T+1)\left(15T^4 - 7T^3 - 12T^2 + 8T + 8\right)] \\
& +\sigma_u^4[\frac{1}{15\,120}(T-2)\left(4037T^5 + 1717T^4 - 4811T^3\right. \\
& \left.+779T^2 + 4758T + 1584\right)], \qquad \text{(A.24)}
\end{aligned}
$$

and for the third term we find that,

$$
\begin{aligned}
& E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)^2 \\
= \; & \mu_4[\sum_{t=2}^{T}(\sum_{k=t}^{T}\sum_{j=t}^{T} jm_{kj})^2] \\
& +\sigma_u^4\{2\sum_{t=2}^{T}[(\sum_{k=t}^{T}\sum_{j=t}^{T} jm_{kj})(\sum_{s=t+1}^{T}(\sum_{k=s}^{T}\sum_{j=s}^{T} jm_{kj}))] + \sum_{t=2}^{T-1}\sum_{s=t+1}^{T}(\sum_{k=t}^{T}\sum_{j=s}^{T} jm_{kj})^2 \\
& +\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}(\sum_{k=s}^{T}\sum_{j=t}^{T} jm_{kj})^2 + 2\sum_{t=2}^{T-1}[\sum_{s=t+1}^{T}(\sum_{k=t}^{T}\sum_{j=s}^{T} jm_{kj})(\sum_{k=s}^{T}\sum_{j=t}^{T} jm_{kj})]\} \\
= \; & \mu_4[\frac{1}{13\,860}(1641T^8 + 726T^7 - 3760T^6 - 594T^5 + 4259T^4 \\
& +792T^3 + 1652T^2 + 1452T - 4728)]/\left(T(T^2-1)\right) \\
& +\sigma_u^4[\frac{1}{69\,300}(T-2)\left(17\,875T^8 + 7362T^7 - 34\,963T^6\right. \\
& +7726T^5 + 50\,465T^4 + 10\,018T^3 + 4603T^2 - 6026T \\
& \left.-35\,460\right)]/\left(T(T^2-1)\right). \qquad \text{(A.25)}
\end{aligned}
$$

Substituting (A.23)-(A.25) for $E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)^2$, $E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)$ $(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)$ and $E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)^2$, respectively, on the right-hand side of (A.22) gives (A.19). Similarly we can show that (A.17), (A.18), (A.20), and (A.21) hold. ■

Lemma 6 *Consider model (3.3) when (3.4) and (A1)-(A3) in Assumption 1 hold. Let $W_{i,jt}$, $j = 1, ..., 5$, be defined in (3.5). Then, for any fixed $T > 1$,*

$$
\begin{aligned}
E(W_{i,1t}W_{i,2t}) &= \frac{1}{83\,160}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\} \times \{\mu_4(155T^6+198T^5 \\
&\quad +104T^4-5677T^2-4950T+7290)+\sigma_u^4[\frac{1}{5}(935T^7 \\
&\quad -1038T^6-3157T^5-570T^4-4675T^3+82\,878T^2 \\
&\quad +78\,177T-109\,350)]\}, \\
E(W_{i,1t}W_{i,3t}) &= \frac{1}{420}\left\{(T^2-4)/\left(T(T-1)\right\}\right. \times \{\mu_4\left(T^2-5\right)\left(T^2+5\right) \\
&\quad +\sigma_u^4[\frac{1}{10}\left(13T^5-30T^4+10T^3-23T+750\right)]\}, \\
E(W_{i,1t}W_{i,4t}) &= -\frac{1}{420}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\} \times \{\mu_4\left(14T^3-15T^2-14T-33\right) \\
&\quad +\sigma_u^4\left(14T^4-42T^3+31T^2+42T+99\right)\}, \\
E(W_{i,1t}W_{i,5t}) &= -\frac{1}{840}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\} \times \{\mu_4(17T^4-15T^3+9T^2 \\
&\quad -33T-170)+\sigma_u^4[\frac{1}{15}(253T^5-555T^4+175T^3-615T^2 \\
&\quad +1732T+7650)]\}, \\
E(W_{i,2t}W_{i,3t}) &= \frac{1}{166\,320}\left\{(T^2-4)/\left(T(T-1)\right)\right\} \times \{\mu_4(245T^6+198T^5 \\
&\quad +290T^4-11\,839T^2-4950T+26\,424) \\
&\quad +\sigma_u^4[\frac{5}{2}(539T^7-1194T^6-649T^5-1680T^4 \\
&\quad -8899T^3+87\,654T^2+44\,649T-198\,180)]\}, \\
E(W_{i,2t}W_{i,4t}) &= -\frac{1}{2520}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\} \times \{\mu_4(30T^5+5T^4+27T^3 \\
&\quad -29T^2-489T-312)+\sigma_u^4[\frac{1}{5}(163T^6-317T^5-95T^4 \\
&\quad -545T^3+292T^2+7342T+4680)]\},
\end{aligned}
$$

$$
\begin{aligned}
E(W_{i,2t}W_{i,5t}) &= -\frac{1}{166\,320}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\}\times\{\mu_4(1682T^6+1089T^5\\
&\quad +1502T^4-396T^3-34\,792T^2-29\,205T+18\,648)\\
&\quad +\sigma_u^4[\frac{1}{5}(6490T^7-8598T^6-11\,825T^5-64\,110T^4-33\,770T^3\\
&\quad +546\,828T^2+466\,785T-279\,720)]\},\\
E(W_{i,3t}W_{i,4t}) &= -\frac{1}{840}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\}\times\{\mu_4(13T^4-T^3+7T^2\\
&\quad -47T-164)+\sigma_u^4[\frac{1}{15}(208T^5-375T^4-5T^3-525T^2\\
&\quad +1957T+7380)]\},\\
E(W_{i,3t}W_{i,5t}) &= -\frac{1}{5040}\left\{(T^2-4)/\left(T(T^2-1)\right)\right\}\times\{\mu_4(63T^5+10T^4\\
&\quad +24T^3-58T^2-951T-624)+\sigma_u^4[\frac{1}{5}(251T^6-319T^5\\
&\quad -280T^4-1720T^3+749T^2+14\,999T+9360)]\},\\
E(W_{i,4t}W_{i,5t}) &= \frac{1}{210}\left\{(T-2)/\left(T(T^2-1)\right)\right\}\times\{\mu_4(29T^4-3T^3-40T^2\\
&\quad -15T-43)+\sigma_u^4[\frac{1}{2}(63T^5-104T^4-52T^3+170T^2\\
&\quad +97T+258)]\}.
\end{aligned}
$$

Proof. We only give the proof for $E(W_{i,2t}W_{i,3t})$. Thus, write

$$
\begin{aligned}
&E(W_{i,2t}W_{i,3t})\\
=\;& E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T^2\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)\\
&-E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T^2\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)\\
&-E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)\\
&+E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i). \qquad \text{(A.26)}
\end{aligned}
$$

Further manipulation for those terms on the right-hand side of (A.26), by applying Lemmas 2 and 4, yields,

$$
\begin{aligned}
&E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T^2\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)\\
=\;& \mu_4\{\sum_{t=1}^{T-1}[(\sum_{j=1}^{T}j^2-\sum_{k=1}^{t}k^2)(\sum_{j=1}^{T}j-\sum_{k=1}^{t}k)]\}\\
&+\sigma_u^4\{\sum_{t=1}^{T-2}[(\sum_{j=1}^{T}j^2-\sum_{k=1}^{t}k^2)(\sum_{j=t+1}^{T-1}(\sum_{s=1}^{T}s-\sum_{k=1}^{j}k))]\\
&+\sum_{t=1}^{T-2}[(\sum_{j=1}^{T}j-\sum_{k=1}^{t}k)(\sum_{j=t+1}^{T-1}(\sum_{s=1}^{T}s^2-\sum_{k=1}^{j}k^2))]
\end{aligned}
$$

$$
\begin{aligned}
&+4\sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}[(\sum_{j=1}^{T}j^2-\sum_{k=1}^{s}k^2)(\sum_{j=1}^{T}j-\sum_{k=1}^{s}k)]\}\\
=\;&\mu_4[\frac{1}{72}T^2\left(7T+2\right)\left(T-1\right)\left(T+1\right)^2]\\
&+\sigma_u^4[\frac{1}{840}T\left(T^2-1\right)\left(T-2\right)\left(178T^3+209T^2+15T-12\right)],
\end{aligned}
\tag{A.27}
$$

$$
\begin{aligned}
&E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T^2\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)\\
=\;&\mu_4\{\sum_{t=1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T}jm_{kj})(\sum_{j=1}^{T}j^2-\sum_{k=1}^{t}k^2)]\}\\
&+\sigma_u^4\{\sum_{t=1}^{T-2}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T}jm_{kj})(\sum_{j=t+1}^{T-1}(\sum_{s=1}^{T}s^2-\sum_{k=1}^{j}k^2))]\\
&+2\sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T}jm_{kj})(\sum_{j=1}^{T}j^2-\sum_{k=1}^{s}k^2)]\\
&+2\sum_{s=1}^{T-2}\sum_{t=s+1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T}jm_{kj})(\sum_{j=1}^{T}j^2-\sum_{k=1}^{t}k^2)]\}\\
=\;&\mu_4[\frac{1}{15\,120}(1367T^6+1635T^5-1051T^4-1263T^3\\
&+1196T^2+1572T+144)]\\
&+\sigma_u^4[\frac{1}{7560}\left(T-2\right)\left(1541T^6+1806T^5-1030T^4\right.\\
&\left.-678T^3+1793T^2+1608T+288\right)],
\end{aligned}
\tag{A.28}
$$

$$
\begin{aligned}
&E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)\\
=\;&\mu_4\{\sum_{t=1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T}kjm_{kj})(\sum_{j=1}^{T}j-\sum_{k=1}^{t}k)]\}\\
&+\sigma_u^4\{\sum_{t=1}^{T-2}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T}kjm_{kj})(\sum_{j=t+1}^{T-1}(\sum_{s=1}^{T}s-\sum_{k=1}^{j}k))]\\
&+\sum_{t=1}^{T-2}[(\sum_{j=1}^{T}j-\sum_{k=1}^{t}k)(\sum_{j=t+1}^{T-1}(\sum_{k=j+1}^{T}\sum_{s=j+1}^{T}ksm_{ks}))]\\
&+4\sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T}kjm_{kj})(\sum_{j=1}^{T}j-\sum_{k=1}^{s}k)]\}
\end{aligned}
$$

$$
\begin{aligned}
= \; & \mu_4[\frac{1}{7560}(691T^6 + 855T^5 - 509T^4 - 783T^3 + 394T^2 + 792T + 144)] \\
& +\sigma_\mu^4[\frac{1}{3780}(T-2)(760T^6 + 879T^5 - 554T^4 \\
& -492T^3 + 694T^2 + 765T + 180)],
\end{aligned} \tag{A.29}
$$

and

$$
\begin{aligned}
& E(\mathbf{u}_i'\mathbf{C}_T'\mathbf{D}_T\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)(\mathbf{u}_i'\mathbf{C}_T'\mathbf{M}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i) \\
= \; & \mu_4\{\sum_{t=1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T} kjm_{kj})(\sum_{j=t+1}^{T}\sum_{k=t+1}^{T} jm_{kj})]\} \\
& +\sigma_u^4\{\sum_{t=1}^{T-2}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T} kjm_{kj})(\sum_{j=t+1}^{T-1}(\sum_{s=j+1}^{T}\sum_{k=j+1}^{T} sm_{ks}))] \\
& +\sum_{t=1}^{T-2}[(\sum_{k=t+1}^{T}\sum_{j=t+1}^{T} jm_{kj})(\sum_{j=t+1}^{T-1}(\sum_{s=j+1}^{T}\sum_{s=j+1}^{T} ksm_{ks}))] \\
& +2\sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T} jm_{kj})(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T} kjm_{kj})] \\
& +2\sum_{s=1}^{T-2}\sum_{t=s+1}^{T-1}[(\sum_{k=t+1}^{T}\sum_{j=s+1}^{T} jm_{kj})(\sum_{k=s+1}^{T}\sum_{j=t+1}^{T} kjm_{kj})]\}
\end{aligned}
$$

$$
\begin{aligned}
= \; & \mu_4[\frac{1}{83\,160T(T-1)}(7157T^8 + 1067T^7 - 13\,952T^6 \\
& +44T^5 + 11\,887T^4 + 1661T^3 + 21\,908T^2 \\
& +7524T - 52\,848)] \\
& +\sigma_\mu^4[\frac{1}{46\,200T(T-1)}(T-2)(8976T^8 + 1356T^7 \\
& -14\,723T^6 + 5768T^5 + 16\,904T^4 + 6014T^3 \\
& +9883T^2 - 16\,058T - 44\,040)].
\end{aligned} \tag{A.30}
$$

By (A.27)-(A.30) further algebra gives (A.26). Similarly, we can show that the remaining formulas in Lemma 6 hold. ■

Lemma 7 *Consider model (3.3) when (3.4) and (A1)-(A3) in Assumption 1 hold. Then, for any fixed $T > 2$,*

$$
\sqrt{n}\left[\begin{pmatrix} \widehat{\rho} - 1 \\ \widehat{\phi} \end{pmatrix} - \begin{pmatrix} B_1(T) \\ B_2(T) \end{pmatrix}\right] \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{Q}_T^{-1}\mathbf{\Omega}_T\mathbf{Q}_T^{-1}\right), \tag{A.31}
$$

where

$$\mathbf{Q}_T = \begin{pmatrix} E\left(W_{i,1t}\right) & E\left(W_{i,3t}\right) \\ E\left(W_{i,3t}\right) & E\left(W_{i,2t}\right) \end{pmatrix}, \tag{A.32}$$

and

$$\mathbf{\Omega}_T = E\begin{bmatrix} E\left(G_{1t}^2\right) & E\left(G_{1t}G_{2t}\right) \\ E\left(G_{1t}G_{2t}\right) & E\left(G_{2t}^2\right) \end{bmatrix}, \tag{A.33}$$

whereas

$$\begin{aligned} G_{i,1t} &= W_{i,4t} - B_1(T)W_{i,1t} - B_2(T)W_{i,3t}, && \text{(A.34)} \\ G_{i,2t} &= W_{i,5t} - B_1(T)W_{i,3t} - B_2(T)W_{i,2t}, && \text{(A.35)} \\ B_2(T) &= -3.5/(T^2-2). && \text{(A.36)} \end{aligned}$$

Proof. (i) The inconsistency of $\widehat{\phi}$ given by $B_2(T)$ in (A.36) is the lower element of the limit vector on the right-hand side of (A.16).

(ii) The LSDV estimator of $(\rho, \phi)'$, under the null hypothesis (3.4), corrected by $[B_1(T),\ B_2(T)]'$, can be expressed as

$$\begin{aligned} &\begin{bmatrix} \widehat{\rho}-1 \\ \widehat{\phi} \end{bmatrix} - \begin{bmatrix} B_1(T) \\ B_2(T) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n W_{i,1t} & \sum_{i=1}^n W_{i,3t} \\ \sum_{i=1}^n W_{i,3t} & \sum_{i=1}^n W_{i,2t} \end{bmatrix}^{-1} \\ &\times \left(\begin{bmatrix} \sum_{i=1}^n W_{i,4t} \\ \sum_{i=1}^n W_{i,5t} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^n W_{i,1t} & \sum_{i=1}^n W_{i,3t} \\ \sum_{i=1}^n W_{i,3t} & \sum_{i=1}^n W_{i,2t} \end{bmatrix} \begin{bmatrix} B_1(T) \\ B_2(T) \end{bmatrix} \right) \\ &= \begin{bmatrix} \sum_{i=1}^n W_{i,1t} & \sum_{i=1}^n W_{i,3t} \\ \sum_{i=1}^n W_{i,3t} & \sum_{i=1}^n W_{i,2t} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n G_{i,1t} \\ \sum_{i=1}^n G_{i,2t} \end{bmatrix}, \end{aligned} \tag{A.37}$$

where $W_{i,jt}$, $j = 1, ..., 5$, are defined in (3.5), $G_{i,1t}$ and $G_{i,2t}$ are given by (A.34) and (A.35), respectively. Note that for a fixed T, $(G_{i,1t}, G_{i,2t})'$ is by Assumption 1 an i.i.d. vector across i with zero mean and finite covariance matrix $\mathbf{\Omega}_T$ defined in (A.33). It follows from Lemmas 5 and 6 that

$$
\begin{aligned}
E\left(G_{i,1t}^2\right) &= E\left(W_{i,4t}^2\right) + B_1^2(T)E\left(W_{i,1t}^2\right) + B_2^2(T)E\left(W_{i,3t}^2\right) \\
&\quad -2B_1(T)E\left(W_{i,4t}W_{i,1t}\right) - 2B_2(T)E\left(W_{i,4t}W_{i,3t}\right) \\
&\quad +2B_1(T)B_2(T)E\left(W_{i,3t}W_{i,1t}\right) \\
&= \mu_4\frac{1}{110\,880}\{[(T-3)(T-2)/((T^2-2)^2(T^2-1)(T+2)T)] \\
&\quad \times(6543T^7 - 13\,599T^6 - 36\,602T^5 + 170\,566T^4 \\
&\quad -25\,173T^3 - 707\,131T^2 - 29\,344T + 589\,812)\} \\
&\quad +\sigma_\mu^4\{\frac{1}{3326\,400}(T-2)/((T^2-2)^2(T^2-1)(T+2)T)\} \\
&\quad \times(392\,843T^9 - 1987\,498T^8 + 3181\,711T^7 + 5981\,726T^6 \\
&\quad -34\,139\,267T^5 + 37\,748\,138T^4 + 74\,407\,229T^3 - 182\,674\,806T^2 \\
&\quad -70\,246\,356T + 159\,249\,240),
\end{aligned}
$$

$$
\begin{aligned}
E\left(G_{i,1t}G_{i,2t}\right) &= E\left(W_{i,4t}W_{i,5t}\right) - B_1(T)E\left(W_{i,3t}W_{i,4t}\right) - B_2(T)E\left(W_{i,2t}W_{i,4t}\right) \\
&\quad -B_1(T)E\left(W_{i,1t}W_{i,5t}\right) + B_1^2(T)E\left(W_{i,1t}W_{i,3t}\right) \\
&\quad +B_1(T)B_2(T)E\left(W_{i,1t}W_{i,2t}\right) - B_2(T)E\left(W_{i,3t}W_{i,5t}\right) \\
&\quad +B_1(T)B_2(T)E\left(W_{i,3t}^2\right) + B_2^2(T)E\left(W_{i,2t}W_{i,3t}\right) \\
&= \mu_4\frac{1}{221\,760}\{(T^2-9)(T-2)/((T^2-2)^2(T^2-1)(T+2)T)\} \\
&\quad \times(3647T^7 + 1535T^6 + 3726T^5 - 10\,450T^4 - 179\,845T^3 \\
&\quad -226\,905T^2 + 219\,080T + 348\,188) \\
&\quad +\sigma_u^4\{\frac{1}{6652\,800}(T-2)/((T^2-2)^2(T+1)(T+2)(T-1)T)\} \\
&\quad \times(382\,943T^{10} - 1478\,555T^9 - 1151\,547T^8 + 11\,553\,117T^7 \\
&\quad -1552\,641T^6 - 12\,256\,869T^5 + 21\,309\,007T^4 - 120\,784\,417T^3 \\
&\quad -220\,115\,082T^2 + 156\,524\,484T + 282\,032\,280),
\end{aligned}
$$

and

$$
\begin{aligned}
E\left(G_{i,2t}^2\right) &= E\left(W_{i,5t}^2\right) + B_1^2(T)E\left(W_{i,3t}^2\right) + B_2^2(T)E\left(W_{i,2t}^2\right) \\
&\quad -2B_1(T)E\left(W_{i,5t}W_{i,3t}\right) - 2B_2(T)E\left(W_{i,2t}W_{i,5t}\right) \\
&\quad +2B_1(T)B_2(T)E\left(W_{i,3t}W_{i,2t}\right)
\end{aligned}
$$

$$
\begin{aligned}
= \; & \mu_4\{[\frac{1}{2882\,880}(T-3)(T-2)/\left(\left(T^2-2\right)^2\left(T^2-1\right)(T+2)\,T\right)\} \\
& \times(55\,489T^9+125\,238T^8+298\,513T^7+975\,550T^6-4756\,977T^5 \\
& -18\,293\,686T^4-2616\,677T^3+34\,056\,714T^2+19\,476\,036T \\
& -6731\,784)+\sigma_u^4\{\frac{1}{259\,459\,200}(T-2)/ \\
& \left(\left(T^2-2\right)^2\left(T^2-1\right)(T+2)\,T\right)\}\times(11\,556\,935T^{11}-40\,537\,573T^{10} \\
& -55\,120\,525T^9+286\,295\,547T^8+63\,649\,665T^7+923\,928\,009T^6 \\
& +843\,787\,177T^5-12\,232\,771\,967T^4-10\,684\,267\,460T^3 \\
& +21\,359\,553\,264T^2+17174\,416\,608T-5452\,745\,040).
\end{aligned}
$$

As $n \to \infty$, with T fixed, the Linderberg-Lèvy CLT implies that the numerator of (A.37) converges at the rate $\sqrt{n}$ to a normal random variable,

$$
n^{-1/2}\sum_{i=1}^{n}\begin{bmatrix} G_{i,1t} \\ G_{i,2t} \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Omega}_T\right). \tag{A.38}
$$

On the other hand, the denominator of (A.37) converges in probability at the rate n such that

$$
n^{-1}\sum_{i=1}^{n}\begin{bmatrix} W_{i,1t} & W_{i,3t} \\ W_{i,3t} & W_{i,2t} \end{bmatrix} \xrightarrow{p} \mathbf{Q}_T. \tag{A.39}
$$

where $\mathbf{Q}_T$ defined in (A.32). It follows from (A.38) and (A.39) that as $n \to \infty$ with T fixed, (A.31) holds. ■

Proof of Theorem 3.2. For $T > 2$, $\mathbf{Q}_T$ defined in Lemma 7 is positive definite. Thus, $\mathbf{Q}_T^{-1}$ exists. Let $\mathbf{Q}_T^{-1} = \begin{pmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{pmatrix}$. It follows from Lemma 7 that $\sqrt{n}(\widehat{\rho}-1-B_1(T)) \xrightarrow{d} N(0, \sigma_{\widehat{\rho}}^2(T,\kappa_4))$ holds with

$$
\sigma_{\widehat{\rho}}^2(T,\kappa_4) = q_{11}^2 E\left(G_{i,1t}^2\right) + 2q_{11}q_{12}E\left(G_{i,1t}G_{i,2t}\right) + q_{12}^2 E\left(G_{i,2t}^2\right). \tag{A.40}
$$

Further manipulation in (A.40) by using the results in Lemma 7 gives the expression of $\sigma_{\widehat{\rho}}^2(T,\kappa_4)$ in Theorem 3.2. ■

Proof of Corollaries 3.3. This is an immediate consequence of Theorem 3.2. ■

Bibliography

ANDERSON, T. W., AND C. HSIAO (1981): "Estimation of dynamic models with error components," *Journal of the American Statistical Association*, 76, 598–606.

BAI, J., AND J. SILVESTRE (2003): "Structural changes, common stochastic trends and unit roots in panel data," Unpublished manuscript.

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

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series," *Journal of the American Statistical Association*, 74, 427–431.

GONZÁLEZ, A. G., T. TERÄSVIRTA, AND D. VAN DIJK (2004): "Panel smmoth transition regression and an application to investment under credit constraints," Stockholm School of Economics, Working paper.

HANSEN, B. E. (1999): "Threshold effects in non-dynamic panels: Estimating, testing, and inference," *Journal of Econometrics*, 93, 345–368.

HARRIS, R. D. F., AND E. TZAVALIS (1999): "Inference for unit roots in dynamic panels where the time dimension is fixed," *Journal of Econometrics*, 91, 201–226.

HSIAO, C. (1986): *Analysis of panel data.* Cambridge University Press.

IM, K. S., AND J. LEE (1999): "LM unit root test with panel data; a test robust to structural changes," Wichita State University, Working paper.

IM, K. S., M. H. PESARAN, AND Y. SHIN (2003): "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115, 53–74.

LEVIN, A., C. LIN, AND C. CHU (2002): "Unit root tests in panel data: Asymptotic and finite-sample properties," *Journal of Econometrics*, 108, 1–24.

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

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

NICKELL, S. (1981): "Biases in dynamic models with fixed effects," *Econometrica*, 49, 1417–1426.

PERRON, P. (1990): "Testing for a unit root in a time series with a changing mean," *Journal of Business and Economic Statistics*, 8, 153–162.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

QUAH, D. (1994): "Exploiting cross section variation for unit root inference in dynamic data," *Economic Letters*, 44, 9–19.

RIPATTI, A., AND P. SAIKKONEN (2001): "Vector autoregressive process with nonlinear time trends in cointegrating relations," *Macroeconomic Dynamics*, 5, 557–597.

SILVESTRE, C. I., T. D. BARRIO-CASTRO, AND E. LOPEZ-BAZO (2001): "Level shifts in a panel data based unit root test. An application to the rate of unemployment," Mimeo, Department of Econometrica, University of Barcelona, Spain.

TERÄSVIRTA, T. (1998): "Modeling economic relationships with smooth transition regressions," Handbook of applied economic statistics, 507-552. New York: Dekker.

TZAVALIS, E. (2002): "Structural breaks and unit roots for short panels," Queen Mary University of London, Working paper No. 459.

Chapter 4

Testing Unit Roots in Nonlinear Dynamic Heterogeneous Panels

4.1 Introduction

Testing the unit root hypothesis in linear panels has received much attention in recent years. One reason for this is that many of the classical unit root tests in univariate settings suffer from low power against near unit root alternatives, and a remedy to this problem is to consider a panel of several univariate time series. For instance, Quah (1994) considers a completely homogeneous linear panel with no cross section specific effects. This approach is, however, quite unrealistic and is likely to yield biased estimators, see e.g. Hsiao (1986). A more general panel is introduced in Levin, Lin, and Chu (2002) because it allows for cross section specific intercepts and time trends, as well as serially correlated residuals over time and heterogeneous variances among cross sections. Both Quah (1994) and Levin, Lin, and Chu (2002) base their tests of a common unit root on the least squares pooled estimator, and inference is enhanced by letting both the number of cross sections and observations over time tend to infinity, implying, of course, a limited practical use. However, Harris and Tzavalis (1999) study the same issues as in Levin, Lin, and Chu (2002) but they instead derive analytical results under the assumption of a traditional panel set-up, i.e. the time dimension is fixed and the number of cross sections is seen as large.

One drawback with all these tests mentioned above is that their alternative hypotheses imply that e.g. all cross sections converge to a long-run equilibrium at the same rate. This is too strong to be held in any interesting empirical cases, as concluded in Maddala and Wu (1999). The problem with a too strong/unrealistic alternative hypothesis is relaxed by Im, Pesaran, and Shin (2003). They assume under the alternative hypothesis that a fraction of the total number of cross sections possesses a linear mean-reversion possibly with cross section specific convergence rates, and that the remaining cross sections are non-stationary. In addition, their testing procedure is fundamentally different because they are averaging individual unit root t-test statistics.

Much of the ongoing research focuses on generalizing the panel unit root tests as such that the tests allow for dependence among cross sections, see for instance Phillips and Sul (2003) and Peseran (2003), but with the linearity property kept. We will stress another issue because evidence against the linearity of the adjustment process in univariate time series has recently been found, see Leybourne, Newbold, and Vougas (1998), Harvey and Mills (2002), and Lanne, Lütkepohl, and Saikkonen (2003), among others. Our concern is therefore that if a cross section is modelled as nonlinear, the conventional linear panel approach yields unit root tests with modest power, as outlined in Chapter 3, and studying a panel is not necessarily a solution to obtaining

satisfactory power. Despite the fact that one obtains more observations, the nonlinearity may be too hard to detect. Our aim is therefore to derive a test of a common unit root in a nonlinear dynamic heterogeneous panel allowing for linear, nonlinear, and non-stationary models under the alternative hypothesis. The panel in this chapter is a generalization of the panel smooth autoregressive introduced in Chapter 3. We are using the testing methodology by Im, Pesaran, and Shin (2003), whereas the testing methodology in Chapter 3 is inspired by Harris and Tzavalis (1999).

The unit root tests in this chapter are very simple to conduct and the limiting distributions are (mostly) the standard normal. We allow for serially correlated errors over time and heterogeneous variances among cross sections. The following dimensions of the panel are considered: (i) The number of cross sections and observations over time are fixed. (ii) The observations over time are fixed and the number of cross sections tends to infinity. (iii) The number of observations over time and cross sections tend to infinity (sequential limits).

The rest of the chapter is organized as follows. In Section 4.2 we present a nonlinear heterogeneous dynamic panel. Testing procedures of a unit root in nonlinear panels are discussed in Section 4.3. The unit root tests are presented in Section 4.4. Asymptotic and finite-sample properties are investigated through Monte Carlo experiments in Section 4.5. Section 4.6 concludes. Thereafter an appendix follows with tables and further simulation results.

4.2 A nonlinear heterogeneous dynamic panel

Consider a sample of n cross sections observed over T time periods (not necessarily the same for each cross section so we allow for unbalanced panels). Suppose that the stochastic process $\{y_{it}\}$ is generated by the first-order panel smooth autoregressive (PSTAR(1)) model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\pi}_{i1} + \mathbf{x}_{it}'\boldsymbol{\pi}_{i2}F_i(t) + u_{it}, \quad i = 1, ..., n, \quad t = 1, ..., T, \tag{4.1}$$

where $\mathbf{x}_{it} = (1, y_{i,t-1})'$, $\boldsymbol{\pi}_{i1} = (\pi_{i10}, \pi_{i11})'$, $\boldsymbol{\pi}_{i2} = (\pi_{i20}, \pi_{i21})'$, and $F_i(t)$ is chosen to be the logistic smooth transition function defined by

$$F_i(t; \gamma_i, c_i) = \frac{1}{1 + \exp[-\gamma_i(t - c_i)]} - 1/2. \tag{4.2}$$

In (4.2), $\gamma_i \in [0, \infty)$ measures the speed of transition over time from one regime to another, and where $c_i \in (0, T)$ gives the point in time where the transition will be symmetric around. By the model specification in (4.1) and (4.2), each cross section unit is modelled as the LSTAR(1) model proposed in Lin and

Teräsvirta (1994) (however, modified since we adjust the transition function by subtracting one half). We especially note that all parameters are defined to be heterogeneous for each cross section, by which it also follows that the model in (4.1) with (4.2) is a generalization of the panel logistic smooth autoregressive (PLSTAR) model in Chapter 3. Furthermore, for $\gamma_i \in [0, \infty)$, the transition function $F_i(t, \gamma_i, c_i)$ is an non-decreasing function in t, and $\gamma_i = 0$ implies, by construction, that $F_i(t; 0, c_i) = 0$ holds for all t, and the resulting model in (4.1) is a linear panel autoregressive (PAR) model with parameter vector $\boldsymbol{\pi}_{i1}$. On the other hand, letting $\gamma_i \to \infty$, $F_i(t; \infty, c_i)$ becomes an indicator function such that $F_i(t; \infty, c_i) = -0.5$ for $0 \leq t < c_i$ and $F_i(t; \infty, c_i) = 0.5$ for $c_i \leq t < T$. It is seen that the PLSTAR(1) model nests the panel threshold autoregressive (PTAR) model with parameter vectors $\boldsymbol{\pi}_{i1} - 0.5\boldsymbol{\pi}_{i2}$ and $\boldsymbol{\pi}_{i1} + 0.5\boldsymbol{\pi}_{i2}$ in regime one ($0 \leq t < c_i$) and two ($c_i \leq t < T$), respectively. Finally, u_{it} is an error term such that for all i and t, $\{u_{it}\}$ defines a sequence of independently distributed random variables with zero means and heterogeneous variances σ_i^2.

4.3 Testing procedures

4.3.1 The test statistic

Testing the unit root hypothesis in the PLSTAR(1) model (4.1) is achieved by imposing the following parameter restrictions

$$H_0 : \pi_{i10} = 0, \quad \pi_{i11} = 1, \quad \gamma_i = 0, \tag{4.3}$$

for all i, and the maintained model for each cross section is therefore

$$y_{it} = y_{i,t-1} + u_{it}, \tag{4.4}$$

i.e. a random walk without drift (this is the most relevant null hypothesis since the PLSTAR(1) model does not contain any time trend). The alternative hypothesis could be described as not H_0, meaning that the alternative hypothesis can be a mixture of LSTAR models (and AR models) and unit root processes. It is, however, important that the total number of LSTAR (and AR models) models, n^*, satisfies $n \geq n^* > 0$ where $n^* = \delta n$ and $\delta \in (0, 1]$, to guarantee consistent tests. Furthermore, as pointed out in Luukkonen, Saikkonen, and Teräsvirta (1988), imposing the restriction $\gamma_i = 0$ leads to identification problem since the parameters π_{i20}, π_{i21}, and c_i are not identified under the null hypothesis. To remedy this problem, we replace the transition function in (4.2) with its first-order Taylor expansion around $\gamma_i = 0$. This approximation is feasible since $F_i(t)$ is twice differentiable in γ_i and the first derivative

evaluated at $\gamma_i = 0$ is non-zero. The approximation yields $F_i(t) \approx 0.25(t - c_i)$ (ignoring the remainder). Substituting for this approximation into (4.1) and collecting terms yields the linearized version of the PLSTAR(1) model

$$y_{it} = \tilde{\mathbf{x}}_{it}'\boldsymbol{\alpha}_i + \tilde{u}_{it}, \tag{4.5}$$

where $\tilde{\mathbf{x}}_{it} = (1, t, y_{i,t-1}, ty_{i,t-1})'$, $\boldsymbol{\alpha}_i = (\alpha_{i1}, ..., \alpha_{i4})'$, and $\tilde{u}_{it}$ is an error term adjusted with respect to the Taylor approximation such that $\tilde{u}_{it} = u_{it}$ holds under the null hypothesis.[1] The transformed null hypothesis of a unit root is now given by

$$H_0^{aux} : \alpha_{i1} = 0, \quad \alpha_{i2} = 0, \quad \alpha_{i3} = 1, \quad \alpha_{i4} = 0, \tag{4.6}$$

for all i. We proceed by examining the single hypothesis $\alpha_{i3} = 1$, by running the regression in (4.5) for each cross section and calculating the t_i statistic given by

$$t_i = \frac{\hat{\alpha}_{i3} - 1}{\hat{\sigma}_{\hat{\alpha}_{i3}}}, \quad i = 1, ..., n, \tag{4.7}$$

where $\hat{\alpha}_{i3}$ denotes the OLS estimator of α_{i3}, $\hat{\sigma}_{\hat{\alpha}_{i3}} = S_i^2 \mathbf{r}_1 \left(\sum_{t=1}^T \tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it}\right)^{-1} \mathbf{r}_1'$, $S_i^2 = \sum_{t=1}^T (y_{it} - \tilde{\mathbf{x}}_{it}'\hat{\boldsymbol{\alpha}}_i)^2 / (T-4)$, and $\mathbf{r}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$. However, to simplify matters in following finite-sample analysis, we focus on a modified t-statistic, denoted t_i^m, which is defined by

$$t_i^m \equiv \frac{\hat{\alpha}_{i3} - 1}{\tilde{\sigma}_{\hat{\alpha}_{i3}}}, \quad i = 1, ..., n, \tag{4.8}$$

where $\tilde{\sigma}^2_{\hat{\alpha}_{i3}}$ is defined as in (4.7) but S^2 is replaced with

$$\tilde{S}_i^2 = \Delta\mathbf{y}_i'\mathbf{M}\Delta\mathbf{y}_i/(T-1),$$

where $\Delta\mathbf{y}_i = (\Delta y_{i1}, ..., \Delta y_{iT})'$, $\mathbf{M} = \mathbf{I}_T - \boldsymbol{\iota}_T\boldsymbol{\iota}_T'/T$ is the within transformation matrix where $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones. It is clear that both S_i^2 and $\tilde{S}_i^2$ converge under the null hypothesis in probability to σ_i^2, and thus, the asymptotic distribution for t_i and t_i^m is the same. However, the finite-samples properties for S_i^2 and $\tilde{S}_i^2$ differ.[2] Moreover, our choice of $\tilde{S}_i^2$ is arbitrary and is based on that $\mathbf{M}$ is fixed (non-stochastic) and generates consistent estimates of σ_i^2. Other options, however not analyzed here, would be $\mathbf{M} = \mathbf{I}_T$ or $\mathbf{M} = \mathbf{I_T} - \mathbf{X}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{X}_T'$ where $\mathbf{X}_T = (\boldsymbol{\iota}_T, \boldsymbol{\tau}_T)$ and $\boldsymbol{\tau}_T = \{1, ..., T\}$.

[1]This holds because the remainder from the approximation equates to zero under the null hypothesis.

[2]Some finite sample properties for S_i^2 and $\tilde{S}_i^2$ are presented in the Appendix A.

Before stating how the information from each cross section through the t_i^m test statistic in (4.8) is used to conduct a test for a common unit root, we establish some fundamental finite-sample properties for t_i^m under the null hypothesis (4.6).

4.3.2 Finite-sample properties of the test statistic under the null hypothesis

Three important finite-sample properties of the t_i^m test statistic under the null hypothesis can be observed. The two first properties concern the invariance with respect to y_{i0} and the heterogeneous variances σ_i^2. The last property is that, for all i, and T sufficiently large, the second moment of t_i^m exists. One way to readily confirm the first two properties is to vectorize the auxiliary model in (4.5) and divide the regressors into stochastic and deterministic matrices according to

$$\mathbf{y}_i = \mathbf{X}_T \boldsymbol{\alpha}_{i1} + \mathbf{Z}_T \boldsymbol{\alpha}_{i2} + \tilde{\mathbf{u}}_i, \tag{4.9}$$

where $\mathbf{X}_T$ is defined as before, $\boldsymbol{\alpha}_{i1} = (\alpha_{i1}, \alpha_{i2})'$, $\mathbf{Z}_T = (\mathbf{y}_{i,-1},\ \mathbf{D}_T \mathbf{y}_{i,-1})$, $\mathbf{y}_{i,-1} = (y_{i0},\ y_{i1}, ..., y_{i,T-1})'$, $\mathbf{D}_T = diag\{\boldsymbol{\tau}_T\}$ is a $T \times T$ matrix with a time trend on its diagonal, $\boldsymbol{\alpha}_{i2} = (\alpha_{i3}, \alpha_{i4})'$, and $\tilde{\mathbf{u}}_i = (\tilde{u}_{i1}, ..., \tilde{u}_{i,T})'$. The partitioned regression in (4.9) implies that we can rewrite (4.8) under the null hypothesis (4.6) as

$$t_i^m = \frac{\mathbf{r}_2 \left[(\mathbf{Q}_T \mathbf{Z}_T)' (\mathbf{Q}_T \mathbf{Z}_T) \right]^{-1} \times \left[(\mathbf{Q}_T \mathbf{Z}_T)' (\mathbf{Q}_T \mathbf{u}_i) \right]}{\tilde{S}_i \sqrt{\mathbf{r}_2 \left[(\mathbf{Q}_T \mathbf{Z}_T)' (\mathbf{Q}_T \mathbf{Z}_T) \right]^{-1} \mathbf{r}_2'}}, \tag{4.10}$$

where $\mathbf{r}_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{X}_T (\mathbf{X}_T' \mathbf{X}_T)^{-1} \mathbf{X}_T'$. Furthermore, under (4.6) we can express $\mathbf{y}_{i,-1}$ as

$$\mathbf{y}_{i,-1} = \boldsymbol{\iota}_T y_{i0} + \mathbf{C}_T \mathbf{u}_i, \tag{4.11}$$

where

$$\mathbf{C}_T = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix}_{T \times T}. \tag{4.12}$$

Using (4.11) we obtain that $\mathbf{Z}_T = (\boldsymbol{\iota}_T y_{i0} + \mathbf{C}_T \mathbf{u}_i,\ \mathbf{D}_T \boldsymbol{\iota}_T y_{i0} + \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)$, and

because $\mathbf{Q}_T$ is orthogonal to both $\boldsymbol{\iota}_T y_{i0}$ and $\mathbf{D}_T\boldsymbol{\iota}_T y_{i0}$, $\mathbf{Q}_T\mathbf{Z}_T = (\mathbf{Q}_T\mathbf{C}_T\mathbf{u}_i, \mathbf{Q}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i)$ holds. Furthermore, under the null hypothesis $\Delta\mathbf{y}_i = \mathbf{u}_i$ which yields $\tilde{S}_i^2 = \mathbf{u}_i'\mathbf{M}\mathbf{u}_i/(T-1)$. It follows, for all i and T, that the t_i^m test statistic is invariant with respect to y_{i0}

To verify the second property, write t_i^m as

$$\begin{aligned} t_i^m &= \sqrt{T-1} \times \frac{\mathbf{r}_2\left[(\mathbf{Q}_T\mathbf{Z}_T/\sigma_i)'(\mathbf{Q}_T\mathbf{Z}_T/\sigma_i)\right]^{-1}}{\sqrt{(\mathbf{M}\mathbf{u}_i/\sigma_i)'(\mathbf{M}\mathbf{u}_i/\sigma_i)}} \\ &\quad \times \frac{\left[(\mathbf{Q}_T\mathbf{Z}_T/\sigma_i)'(\mathbf{Q}_T\mathbf{u}_i/\sigma_i)\right]}{\sqrt{\mathbf{r}_2\left[(\mathbf{Q}_T\mathbf{Z}_T/\sigma_i)'(\mathbf{Q}_T\mathbf{Z}_T/\sigma_i)\right]^{-1}\mathbf{r}_2'}}. \end{aligned} \tag{4.13}$$

Obviously the σ_i's cancel out in (4.13), and in the matrix $\mathbf{Q}_T\mathbf{Z}_T/\sigma_i = (\mathbf{Q}_T\mathbf{C}_T\mathbf{u}_i/\sigma_i, \mathbf{Q}_T\mathbf{D}_T\mathbf{C}_T\mathbf{u}_i/\sigma_i)$ we see that the sub-vectors $\mathbf{u}_i/\sigma_i$ are identically distributed with mean $E[\mathbf{u}_i/\sigma_i] = \mathbf{0}$ and covariance matrix $E[\mathbf{u}_i\mathbf{u}_i'/\sigma_i^2] = \mathbf{I}_T$. We conclude that t_i^m is invariant with respect to the nuisance parameter σ_i.

The third property, i.e. that $E\left(t_i^m\right)^2$ exists, is confirmed by simulations. It appears that $E\left(t_i^m\right)^2$ exists for $T \in [4, \infty]$, also indicating that the second moment is existing in an asymptotic sense. In a related context Im, Pesaran, and Shin (2003) also rely upon simulations to conclude that the second moment of their test statistic exists. The present case is somewhat more complicated and the proof of this is left for further research. The methods in Larsson (1997) and Nabeya (1999) might be applicable for $T \in [4, \infty)$ and $T \to \infty$, respectively.

The finite-sample properties for the t_i^m tests statistic summarized above combined with that t_i^m defines a measurable function of the i.i.d. sequence $\{\mathbf{u}_i/\sigma_i\}$, are sufficient conditions for us to conduct tests of a common unit root in the nonlinear heterogeneous panel in (4.1).

4.4 The unit root tests

By imposing certain restrictions on n and T we obtain many interesting testing situations in the nonlinear dynamic heterogeneous panel described above.

4.4.1 Fixed T unit root test in a balanced panel letting n tend to infinity

Assumption 1 *Let $\{u_{it}\}$ be a sequence of independently and normally distributed random variables such that, for all i and t, $E[u_{it}] = 0$ and $E[u_{it}^2] = \sigma_i^2 \in \mathbb{R}_{++}$ hold.*

Assumption 2 *Let T be the same for all i (a balanced panel).*

Proposition 4.1 *Under Assumptions 1 and 2, the null hypothesis (4.6), and $3 < T < \infty$, the individual statistics, t_i^m, are i.i.d. with $E[t_i^m] = \mu(T) \in \mathbb{R}$ and $V[t_i^m] = \eta^2(T) \in \mathbb{R}_{++}$. Furthermore, define the random variable $\mathcal{Z}_0 \equiv n^{-1/2} \sum_{i=1}^{N} (t_i^m - \mu(T)) / \eta(T)$. Then, by letting n tend to infinity, the Lindberg-Lèvy central limit theorem gives*

$$\mathcal{Z}_0 \xrightarrow{d}_n N(0,1). \tag{4.14}$$

where $\xrightarrow{d}_n$ denotes convergence in distribution by letting $n \to \infty$.

In Proposition 4.1 it is clear that we use the information from each cross section unit by averaging the t_i^m test statistics, inspired by Im, Pesaran, and Shin (2003). This procedure is in contrast to the one in Chapter 3 where data is pooled and a test for a common unit root is based on the LSDV estimator. It should be pointed out that both the mean $E[t_i^m]$ and the variance $V[t_i^m]$ are under the null hypothesis functions of T and are tabulated in Table A.1 in the Appendix A for different values of T, where it also is seen that $\lim_{T\to\infty} \mu(T) = \mu^* \in \mathbb{R}$ and $\lim_{T\to\infty} \eta^2(T) = \eta^{2*} \in \mathbb{R}_{++}$.

4.4.2 Fixed T unit root test in an unbalanced panel letting n tend to infinity

Assumption 3 *Let T be different for at least one i (an unbalanced panel).*

Proposition 4.2 *Under Assumptions 1 and 3, the null hypothesis (4.6), and $3 < T < \infty$, the individual statistics, t_i^m, are independently heterogeneously distributed with $E[t_i^m] = \mu_i(T) \in \mathbb{R}$, $V[t_i^m] = \eta_i^2(T) \in \mathbb{R}_{++}$, and for any i, $E[(t_i^m)^3] < \infty$, holds. Furthermore, define the random variable $\mathcal{Z}_1 \equiv n^{-1/2} (\bar{t} - \bar{\mu}(T)) / (\bar{\eta}(T))$ where $\bar{t} = n^{-1} \sum_{i=1}^{n} t_i^m$, $\bar{\mu}(T) = n^{-1} \sum_{i=1}^{n} \mu_i(T)$, and $\bar{\eta}(T) = n^{-1} \sum_{i=1}^{n} \eta_i^2(T)$. Then, by letting n tend to infinity, the Liapounov central limit theorem gives*

$$\mathcal{Z}_1 \xrightarrow{d}_n N(0,1). \tag{4.15}$$

In Proposition 4.2 we must assert that $E[(t_i^m)^3] < \infty$ which holds as long as $3 < T$, and is confirmed by the simulation results reported in Table A.1 in the Appendix A

4.4.3 Fixed T and n unit root test in a balanced panel

This is perhaps the most relevant case from a practitioner point of view. However, viewing the number of countries as finite renders the asymptotic inference concluded in Propositions 4.1 and 4.2. For both a fixed T and n, the analytical expression for the finite-sample distribution of $n^{-1}\sum_{i=1}^{n} t_i^m$ is hardly known. However, in Subsection 4.3.2 it is shown that the t_i^m test statistic is nuisance parameter free under the null hypothesis. This means that the finite-sample distributions for $n^{-1}\sum_{i=1}^{n} t_i^m$ is readily obtained by Monte Carlo simulations for any combinations of T and n. These simulations are reported in the next section, and they also provide a measure of how well the n finite-sample distributions approximate the asymptotic $N(0,1)$ distribution.

4.4.4 Asymptotic T and n unit root test

We apply the method of sequential limits, see Phillips and Moon (1999).[3] Specifically, we consider first a fixed cross section unit and let T tend to infinity which yields intermediate asymptotic results, and thereafter we let n tend to infinity. The intermediate univariate asymptotic results that we need follow from Lemma 1 in Chapter 1.

Proposition 4.3 *If Assumption 1 and the null hypothesis in (4.6) hold, then for any i and letting T tend to infinity it follows that*

$$t_i^m \xrightarrow{d}_T \xi_i = \frac{\mathbf{r}_1 \mathbf{\Psi}_{1i}^{-1} \mathbf{\Pi}_{1i}}{\left(\mathbf{r}_1 \mathbf{\Psi}_{1i}^{-1} \mathbf{r}_1'\right)^{1/2}}, \tag{4.16}$$

where $\xrightarrow{d}_T$ denotes convergence in distribution by letting $T \to \infty$, and

$$\mathbf{\Psi}_{1i} = \begin{bmatrix} \mathbf{M}_{11i} & \mathbf{M}_{12i} \\ \mathbf{M}_{12i} & \mathbf{M}_{13i} \end{bmatrix}, \quad \mathbf{\Pi}_{1i} = \begin{bmatrix} \mathbf{\Pi}_{11i} \\ \mathbf{\Pi}_{12i} \end{bmatrix},$$

with sub-matrices given by

$$\mathbf{M}_{11i} = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix},$$

[3] Lemma 6 in Phillips and Moon (1999) might be applicable to find conditions for sequential convergence to imply joint convergence.

$$\begin{aligned}
\mathbf{M}_{12i} &= \begin{bmatrix} \int_0^1 B_i(r)dr & \int_0^1 rB_i(r)dr \\ \int_0^1 rB_i(r)dr & \int_0^1 r^2B_i(r)dr \end{bmatrix}, \\
\mathbf{M}_{13i} &= \begin{bmatrix} \int_0^1 B_i^2(r)dr & \int_0^1 rB_i^2(r) \\ \int_0^1 rB_i^2(r)dr & \int_0^1 r^2B_i^2(r)dr \end{bmatrix}, \\
\mathbf{\Pi}_{11i} &= \begin{bmatrix} B_i(1) \\ B_i(1) - \int_0^1 B_i(r)dr \end{bmatrix}, \\
\mathbf{\Pi}_{12i} &= \begin{bmatrix} 0.5\left(B_i(1)^2 - 1\right) \\ 0.5\left(B_i(1)^2 - \int_0^1 B_i^2(r)dr - 1/2\right) \end{bmatrix},
\end{aligned}$$

where $B_i(r)$ denotes a standard Brownian motion with respect to u_i on $[0,1]$. Assume further that, $B_i(r)$ and $B_j(r)$ are independent for $i \neq j$. For $i = 1, ..., n$, the limiting distributions ξ_i are i.i.d. with $E[\xi_i] = \mu^* \in \mathbb{R}$ and $V[\xi_i] = \eta^{*2} \in \mathbb{R}_{++}$. Furthermore, define the random variable $\mathcal{Z}_2 \equiv n^{-1/2}\sum_{i=1}^n (\xi_i - \mu^*)/\eta^*$. Then by letting n tend to infinity, the Lindberg-Lèvy central limit theorem gives,

$$\mathcal{Z}_2 \xrightarrow{d}_{n,T} N(0,1), \tag{4.17}$$

where $\xrightarrow{d}_{n,T}$ denotes convergence in distribution by first letting $T \to \infty$ followed by $n \to \infty$.

4.4.5 Asymptotic T and n unit root test in a panel with serially correlated errors

From the results in Chapter 2 we note that a unit root test that accommodates serially correlated errors over time can easily be imposed by adding the terms $\triangle y_{i,t-j}$ $(j \geq 1)$ to the auxiliary regression model in (4.5). This principle is the analogy to the classical ADF tests. We obtain

$$y_{it} = \tilde{\mathbf{x}}_{it}'\boldsymbol{\alpha}_i^a + \triangle\mathbf{y}_{it}'\boldsymbol{\zeta}_i + u_{it}, \tag{4.18}$$

where $\tilde{\mathbf{x}}_{it}$ is the same vector of explanatory variables as in (4.5), $\boldsymbol{\alpha}_i^a = (\beta_i, \delta_i, \rho_i, \psi_i)'$, $\triangle\mathbf{y}_{it} = (\triangle y_{i,t-1},\ \triangle y_{i,t-2}, ..., \triangle y_{i,t-p_i+1})'$, $p_i \geq 2$ denotes the order of augmentation for cross section i, $\boldsymbol{\zeta}_i = (\zeta_{i1}, \zeta_{i2}, ..., \zeta_{i,t-p_i-1})'$, and u_{it} is an error term that fulfills Assumption 1.[4]

Assumption 4 *Assume that, for all i, the roots of the characteristic polynomial $\phi(z) = 1 - \zeta_{i1}z - \zeta_{i2}z^2 - ... - \zeta_{i,p_i-1}z^{p_i-1}$ lie outside the unit circle.*

[4]The regression in (4.18) corresponds to the NPADF testing equations in Chapter 2.

Assumption 4 rules out the possibility of a cross section unit being integrated of order two. The auxiliary null hypothesis of a single unit root for each cross section, under Assumption 4, can now be formulated as

$$H_0^{aux} : \beta_i = 0, \quad \delta_i = 0, \quad \rho_i = 1, \quad \psi_i = 0, \quad \boldsymbol{\zeta}_i \in \mathbb{R}^{p_i - 1}, \tag{4.19}$$

for all i. The alternative hypothesis is not H_0^{aux}, now meaning a mixture of LSTAR models (and AR models) and unit root processes with serially correlated errors, though still in such proportions that the test remains consistent. Under Assumption 4 and the null hypothesis (4.19) the following data generating process (DGP) is obtained

$$y_{it} = y_{i,t-1} + \zeta_{i1}\Delta y_{i,t-1} + \zeta_{i2}\Delta y_{i,t-2} + ... + \zeta_{i,p_i-1}\Delta y_{i,t-p_i+1} + u_{it}. \tag{4.20}$$

Rewrite (4.20) as

$$\Delta y_t = \varepsilon_{it}, \tag{4.21}$$

where $\varepsilon_{it} = (1 - \zeta_{i1}L - \zeta_{i2}L^2 - ... - \zeta_{i,p_i-1}L^{p_i-1})^{-1}u_{it} = \Psi(L)u_{it}$, where L denotes the lag operator and $\Psi(L)$ is a one-sided moving average polynomial in the lag operator. The process in (4.21) clearly defines a stochastic process with serially correlated increments. Moreover, proceed by running the regression in (4.18) to calculate, for each i, the augmented t_i statistic, denoted t_i^a, by

$$t_i^a \equiv \frac{\hat{\rho}_i - 1}{\hat{\sigma}_{\hat{\rho}_i}}, \quad i = 1, ..., n, \tag{4.22}$$

where $\hat{\rho}_i$ denotes the OLS estimator of ρ_i, $\hat{\sigma}_{\hat{\rho}_i} = S_{ia}^2 \left(\mathbf{r}_3 \left[\sum_{t=1}^T \tilde{\mathbf{X}}_{it}' \tilde{\mathbf{X}}_{it} \right]^{-1} \mathbf{r}_3' \right)$, $S_{ia}^2 = \sum_{t=1}^T \left(y_{it} - \tilde{\mathbf{x}}_{it}' \hat{\boldsymbol{\alpha}}_{i1}^a - \Delta \mathbf{y}_{it}' \hat{\boldsymbol{\zeta}}_i \right)^2 / (T - 4 - p_i + 1)$, $\mathbf{r}_3 = \begin{bmatrix} \mathbf{r}_1' & \mathbf{0}_{1\times(p_i-1)}' \end{bmatrix}$, and $\tilde{\mathbf{X}}_{it} = \begin{bmatrix} \tilde{\mathbf{x}}_{it}' & \Delta \mathbf{y}_{it}' \end{bmatrix}'$. Notice that in the case of serial correlation and for a finite T, one encounters the problem with nuisance parameters. For a finite T and any i, the t_i^a test statistic is dependent on σ_i^2, $\{\zeta_{ij}\}_{j=1}^{p_i-1}$, p_i, and the starting values $\mathbf{y}_{i0} = (y_{i,-p_i+1}, ..., y_{i0})'$. However, if one assume that $\mathbf{y}_{i0} = \mathbf{0}$, then t_i^a still depends upon $\{\zeta_{ij}\}_{j=1}^{p_i-1}$ and p_i, but the invariance with respect to σ_i^2 is resurrected.[5] As a result, letting $\mathbf{y}_{i0} = \mathbf{0}$, implies that the expected

[5]For an example, let $p_i = 2$ in (4.18), and pre-multiply the matrix of stochastic regressors implied by (4.18) $[\boldsymbol{\iota}_T y_{i0} + \mathbf{C}_T \mathbf{u}_i, \mathbf{D}_T \boldsymbol{\iota}_T y_{i0} + \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i, \mathbf{c}_0 + \mathbf{C}_T \mathbf{u}_i]$ where $\mathbf{c}_0 = (y_{i0} - y_{i,-1}) \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}'$ with the $\mathbf{Q}_T$ matrix. Notice that $\mathbf{Q}_T$ is orthogonal to $\boldsymbol{\iota}_T y_{i0}$ and $\mathbf{D}_T \boldsymbol{\iota}_T y_{i0}$ but not to $\mathbf{c}_0$. This indicates that t_i^a is dependent on the two starting values y_{i0} and $y_{i,-1}$. Furthermore, recall that σ_i cancels out in (4.13), but in the the present case division with σ_i yields expressions on the form $\mathbf{c}_0/\sigma_i$ that will not further simplify, and σ_i will affect the test statistic through the starting values.

value and the variance of t_i^a only will depend on p_i, $\{\zeta_{ij}\}_{j=1}^{p_i-1}$, and T, and which is denoted $E[t_i^a(T, p_i, \zeta_i)]$ and $V[t_i^a(T, p_i, \zeta_i)]$ respectively. In Appendix A these expected values and variances are tabulated for different values of T, p_i, and ζ_i. However, the factors that t_i^a is dependent upon in finite-samples are eliminated if one let $T \to \infty$. We conclude the following important result.

Proposition 4.4 *If Assumptions 1 and 4 and the null hypothesis (4.19) hold, then, for any i and letting T tend to infinity, it follows that*

$$t_i^a \xrightarrow{d}_T \xi_i, \tag{4.23}$$

where ξ_i is the same limiting distribution as in Proposition 4.3. Therefore, by defining the random variable $\mathcal{Z}_3 \equiv n^{-1/2} \sum_{i=1}^{n} (\xi_i - \mu^*) / \eta^*$ and letting n tend to infinity, we obtain the same result as in (4.17).

Proof. The proof is similar to the proof of Corollary 2.7 in Chapter 2, and is therefore omitted. ■

4.5 Monte Carlo experiments

In this section we examine the finite-sample properties of the $\mathcal{Z}_0$ and $\mathcal{Z}_3$ test statistics in Proposition 4.1 and Proposition 4.4 respectively, by Monte Carlo simulations . The two first Monte Carlo experiments assess the size properties of the tests. In the remaining Monte Carlo experiments, the empirical power of the tests are examined. All experiments are carried under the assumption of balanced panels.

4.5.1 Size properties

4.5.1.1 Estimated size in the case of no serial correlation

The first Monte Carlo experiment examines the $\mathcal{Z}_0$ test statistic (no serial correlation) in Proposition 4.1 and its size properties when the DGP is given by

$$y_{it} = y_{i,t-1} + u_{it}, \quad t = 1, ..., T, \quad i = 1, ..., n. \tag{4.24}$$

In (4.24), it is assumed that $u_{it} \sim nid(0, \sigma_i^2)$ and $\sigma_i^2 \sim U[0.5, 1.5]$ for all i and where U denotes the uniform distribution. The size results are shown in Table 4.1.

We see in Table 4.1 that the size distortions for all T and n of the $\mathcal{Z}_0$ test are negligible at both 5% and 10% significance levels. We conclude that

Table 4.1 The size of the $\mathcal{Z}_0$ test in Proposition 4.1. No serial correlation.

n \ T	25	50	100	250
5	0.051 (0.101)	0.050 (0.100)	0.049 (0.101)	0.050 (0.099)
25	0.049 (0.099)	0.050 (0.101)	0.049 (0.100)	0.050 (0.100)
50	0.048 (0.098)	0.051 (0.100)	0.051 (0.100)	0.050 (0.101)

Notes: The nominal sizes of the test are 5% and 10% (in parentheses). The results are based on 10 000 replications.

the convergence to the asymptotic standard normal distribution is very fast. Already for n as small as five, the approximation to the standard normal distribution is excellent.

4.5.1.2 Estimated size in the case of serially correlated errors

In the second Monte Carlo experiment we examine the size properties of the $\mathcal{Z}_3$ test statistic (serially correlated errors) in Proposition 4.4 when u_{it} in (4.24) is replaced by the AR(1) process, $u_{it} = \rho u_{it} + v_{it}$, where $\rho \in \{0.3, 0.6\}$ and $v_{it} \sim nid(0,1)$ for all i and t. These results are presented in Table 4.2.

In Table 4.2 with $p_i = 0$, the $\mathcal{Z}_3$ test is the same test as $\mathcal{Z}_0$, and is undersized for all values of T, n, and ρ considered. This is to be expected since a test statistic which does not take the serial correlation into account is used. With $p_i = 1$, a first attempt to adjust for the serial correlation, we see that the size distortions are reduced for all T, but increase slowly with n. This supports the well-known fact that over-fitting (the case $p_i \geq 1$), is less harmful than under-fitting (the case $p_i < 1$). We also see that the distortions are more severe when ρ is increased. Moreover, for $T > 50$, $p_i \geq 1$, and all n, the size distortions are modest, however always larger for a more persistent autocorrelation, and that the size distortions are more evident when n is large relatively to T. To this end it is noticed that the remarks about the size for the $\mathcal{Z}_3$ test are in line with the findings for the the IPS test statistic reported in Im, Pesaran, and Shin (2003).

4.5.2 Empirical power

4.5.2.1 A heterogeneous nonlinear panel

In the third Monte Carlo study we examine the empirical power of the $\mathcal{Z}_0$ test statistic in Proposition 4.1 when the DGP accommodates smooth heterogeneous shift in levels and dynamics and is given by

Table 4.2 The size of the $\mathcal{Z}_3$ test in Proposition 4.4. Serial correlation.

			$T = 25$			
n	ρ	p_i	0	1	2	3
5	0.3		0.01 (0.02)	0.05 (0.09)	0.05 (0.10)	0.06 (0.12)
	0.6		0.00 (0.00)	0.05 (0.09)	0.04 (0.10)	0.05 (0.11)
25	0.3		0.00 (0.00)	0.04 (0.07)	0.04 (0.08)	0.06 (0.11)
	0.6		0.00 (0.00)	0.03 (0.06)	0.02 (0.05)	0.02 (0.05)
50	0.3		0.00 (0.00)	0.03 (0.06)	0.03 (0.07)	0.06 (0.11)
	0.6		0.00 (0.00)	0.02 (0.03)	0.01 (0.03)	0.01 (0.03)
			$T = 50$			
n	ρ	p_i	0	1	2	3
5	0.3		0.00 (0.00)	0.05 (0.10)	0.05 (0.11)	0.06 (0.12)
	0.6		0.00 (0.00)	0.05 (0.10)	0.05 (0.10)	0.05 (0.10)
25	0.3		0.00 (0.00)	0.04 (0.09)	0.05 (0.09)	0.06 (0.12)
	0.6		0.00 (0.00)	0.03 (0.07)	0.03 (0.07)	0.03 (0.06)
50	0.3		0.00 (0.00)	0.04 (0.08)	0.05 (0.10)	0.06 (0.13)
	0.6		0.00 (0.00)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)
			$T = 100$			
n	ρ	p_i	0	1	2	3
5	0.3		0.00 (0.01)	0.05 (0.10)	0.05 (0.10)	0.05 (0.10)
	0.6		0.00 (0.00)	0.05 (0.10)	0.05 (0.10)	0.05 (0.10)
25	0.3		0.00 (0.00)	0.04 (0.09)	0.05 (0.10)	0.05 (0.09)
	0.6		0.00 (0.00)	0.04 (0.07)	0.04 (0.08)	0.04 (0.08)
50	0.3		0.00 (0.00)	0.04 (0.08)	0.05 (0.10)	0.04 (0.08)
	0.6		0.00 (0.00)	0.03 (0.06)	0.03 (0.07)	0.03 (0.07)
			$T = 250$			
n	ρ	p_i	0	1	2	3
5	0.3		0.00 (0.01)	0.04 (0.09)	0.05 (0.10)	0.05 (0.10)
	0.6		0.00 (0.01)	0.05 (0.10)	0.05 (0.10)	0.05 (0.10)
25	0.3		0.00 (0.00)	0.04 (0.08)	0.04 (0.09)	0.05 (0.10)
	0.6		0.00 (0.00)	0.04 (0.09)	0.05 (0.09)	0.04 (0.09)
50	0.3		0.00 (0.00)	0.04 (0.08)	0.04 (0.09)	0.05 (0.10)
	0.6		0.00 (0.00)	0.04 (0.08)	0.04 (0.09)	0.04 (0.09)

Notes: The nominal sizes of the test are 5% and 10% (in parentheses). The results are based on 10 000 replications.

$$y_{it} = \pi_{i10} + \pi_{i11} y_{i,t-1} + (\pi_{i20} + \pi_{i21} y_{i,t-1})\, \tilde{F}_i(t) + u_{it}, \tag{4.25}$$

and where we have replaced, without loss of generality, the transition function in (4.2) with $\tilde{F}_i(t) = 1/(1 + \exp[-\gamma_i(t - c_i)])$. This yields more convenient interpretation of the parameters because $\tilde{F}_i(t)$ has the range $[0, 1]$. The parameters in the PLSTAR(1) in (4.25) are assigned the following values

$$\begin{aligned} &\pi_{i10} = 0, && \pi_{i11} \sim U[0.35, 0.45], && \gamma_i \sim U[0.5, 1.5], \\ &c_i \sim U[0.4T, 0.6T], && \pi_{i20} \sim U[0.5, 1.5], && \pi_{i21} \sim U[0.4, 0.5], \\ &u_{it} \sim nid(0, 1). \end{aligned} \tag{4.26}$$

The choice of parameter values in (4.26) implies that all the cross section units display time series that start from the same level because $\pi_{i10} = 0$ holds for all i. The level of a new long-run cross section specific equilibrium is given by $\pi_{i20}/(1-\pi_{i11}-\pi_{i21}) \in [2, 30]$ (assuming that a complete transition takes place). The parameters π_{i20} are set to vary modestly because it is well known that a test based on a first-order Taylor approximation is not designed to capture changes in the intercept, see Luukkonen, Saikkonen, and Teräsvirta (1988) and the discussion in Chapter 1 about the level leverage effect. The panel autoregressive parameters in the linear part in the PLSTAR(1) model yield PAR(1) processes that are modestly persistent because $\pi_{i11} \sim U[0.35, 0.45]$. The panel autoregressive parameters in the nonlinear part of the PLSTAR(1) model are chosen such that $\max_i \pi_{i11} + \max_i \pi_{i21} = 0.95$ (< 1) and the trajectories are therefore stable around the new long-run equilibrium.[6] The speed of transition between regimes varies in the cross sections from 0.5 to 1.5, and implies that a complete transition takes place for all sample sizes that will be considered. The timing of the transitions varies around the mid-point of the sample, i.e. $c_i = 0.5T$, which for instance illustrates that some cross sections respond earlier to, say, a shock and the responses for the other cross sections are somewhat delayed.

The empirical power for the $\mathcal{Z}_0$ test statistic in Proposition 4.1 is presented in Table 4.3. For illustration we also compare the power of our test to the power of the corresponding t-test statistic in Im, Pesaran, and Shin (2003), henceforth abbreviated to the IPS test. It should be noted that we compare

[6] We use the word stable rather than stationary, since the transition function is a function of time. Imposing the restriction $\max_i \pi_{i11} + \max_i \pi_{i21} = 0.95 < 1$ rules out the case of a unit root at the end of the sample period.

Table 4.3 Empirical power of the $\mathcal{Z}_0$ test in Proposition 4.1 and the IPS test. The DGP is a PLSTAR(1) model. No serial correlation.

T	25		50		100	
n	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS
5	0.47	0.00	0.63	0.00	0.87	0.10
25	0.96	0.00	0.99	0.00	1.00	0.15
50	1.00	0.00	1.00	0.00	1.00	0.21

Note: The nominal size is 5%, and the results are based on 10 000 replications.

the power to the IPS test that is based on a linear panel without a time trend.[7] This is reasoned by the fact that our panel lacks the property of a time trend. We also note that the same null hypothesis is tested.

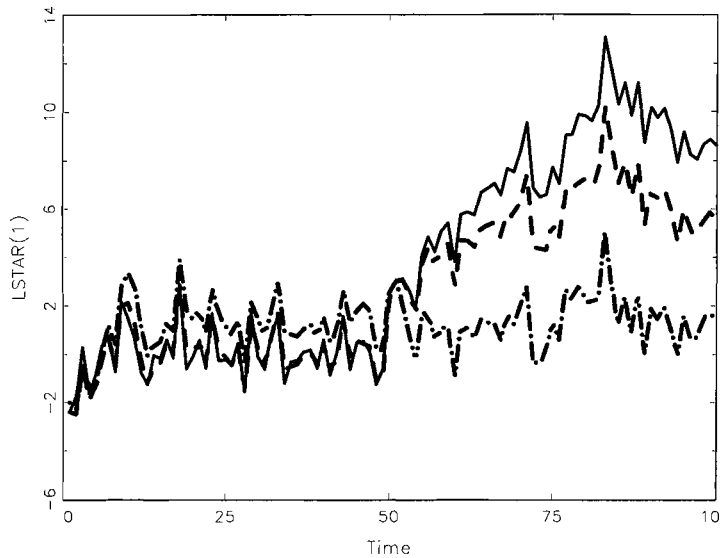
In Table 4.3 we see that the power for the IPS test is rather poor for all T and n. Even for as large panels as $T = 100$ and $n = 50$, the power is only 0.212. The reason for the poor performance of the IPS test is that it is not designed to have power against nonlinear alternatives with a shift in levels. However, as indicated in Chapter 2, the classical univariate Dickey-Fuller t-type of test also has poor power results (close to zero or equal to zero) in a similar set-up to (4.26). In our case this illuminates that a panel approach does not resurrect the power for the Dickey-Fuller t-type of test in the univariate context to satisfactory levels. On the other hand, the $\mathcal{Z}_0$ test performs very satisfactorily and the power is close to unity for $n \geq 25$ and all samples sizes. As a final remark, the power results for the $\mathcal{Z}_0$ test in Table 4.3 is lower than for the test statistic used in Chapter 3 under a similar Monte Carlo experiment (cf. Table 3.3 in Chapter 3). This might be explained by that the test statistic used in Chapter 3 is of deviation type (the deviation form of the LSDV estimator) and not of t-type, and is based on pooled data. These two facts contributes to that the power for the $\mathcal{Z}_0$ test is lower. A typical trajectory generated by the set-up in (4.26) is depicted in Figure 4.1.

4.5.2.2 A nearly linear heterogeneous panel

In the fourth Monte Carlo experiment we investigate the empirical power properties for $\mathcal{Z}_0$ when $\gamma_i = 0.01$ for all i in (4.26), and the resulting model in

[7]The IPS test that we are using averages the classical Dickey-Fuller t-type of test based on an AR(1) process without a time trend.

Figure 4.1 Typical realizations of the LSTAR(1) models in the third (dashed line), fourth (dot-dashed line), and fifth (solid line) Monte Carlo experiment.



(4.25) is nearly linear. The set-up for the remaining parameters are the same as in (4.26). The results are presented in Table 4.4.

From Table 4.4 we conclude that the power of the IPS test is superior, and is expected because the DGP is almost linear. Notable is, however, that the $\mathcal{Z}_0$ test statistic performs satisfactorily and for $n \geq 25$ and $T \geq 50$ the power equals unity. A nearly linear LSTAR(1) model is depicted in Figure 4.1.

4.5.2.3 A homogeneous nonlinear panel

In the fifth Monte Carlo experiment we abandon the randomness of the parameters in (4.25), and all cross sections are represented by the same LSTAR(1) model, i.e. a nonlinear homogeneous panel, with parameters

$$\pi_{i10} = 0, \quad \pi_{i11} = 0.1, \quad \gamma_i = 1, \quad c_i = 0.5T, \quad \pi_{i20} = 1, \quad \pi_{i21} = 0.8, \quad (4.27)$$

for all i. This scenario facilitates the comparison the comparison to the power of a similar univariate unit root test (parameter constancy test) in Chapter 1

Table 4.4 Empirical power of the $\mathcal{Z}_0$ test in Proposition 4.1 and the IPS test. The DGP is an almost linear PLSTAR(1) model. No serial correlation.

T	25		50		100	
n	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS
5	0.12	0.96	0.46	1.00	1.00	1.00
25	0.46	1.00	1.00	1.00	1.00	1.00
50	0.76	1.00	1.00	1.00	1.00	1.00

Note: The nominal size is 5%, and the results are based on 10 000 replications.

Table 4.5 Empirical power of the $\mathcal{Z}_0$ test in Proposition 4.1 and the IPS test. A nonlinear homogeneous panel. No serial correlation.

T	25		50		100	
n	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS
1*	0.11	0.00	0.20	0.00	0.28	0.00
5	0.67	0.00	0.76	0.00	0.90	0.00
25	1.00	0.00	1.00	0.00	1.00	0.00
50	1.00	0.00	1.00	0.00	1.00	0.01

Notes: The nominal size is 5%, and the results are based on 10 000 replications. The asterisk indicates that the power is calculated using the corresponding t-type of tests in Chapter 1 and in Dickey and Fuller (1979).

with the same DGP as in (4.25) and parameter values given by (4.27).[8] The outcome of the experiment is presented in Table 4.5.

In Table 4.5 the benefits by using the panel unit root test statistic $\mathcal{Z}_0$ over a univariate t-type of tests are revealed. For an example, with $n = 1$ and $T = 100$ we see that the power is 0.278, and for $n = 5$ and $T = 100$ the power has increased to 0.895. A small increase in the number of cross sections leads to a substantial improvement in the power. Notable is also that the power for the IPS test is zero for all combinations of T and n. This should be compared to the results for the IPS test in Table 4.3, where the IPS showed some power, e.g. 0.212 for $T = 100$ and $n = 50$. The reason for the reduction in power for the IPS test is that the model in (4.27) generates time series with a more evident change in levels than the time series generated under the third Monte

[8] In Chapter 1 we consider a univariate test of the form $T(\hat{\alpha}_3 - 1)$. The comparison presented here is based on the t-test in (4.7) with $n = 1$.

Table 4.6 Empirical power of the $\mathcal{Z}_3$ test in Proposition 4.4. The DGP is a PLSTAR(1) model. Serial correlation.

T	25				50				100			
		$\mathcal{Z}_3$				$\mathcal{Z}_3$				$\mathcal{Z}_3$		
n	p_i	1	2	3	p_i	1	2	3	p_i	1	2	3
5		0.11	0.10	0.09		0.25	0.21	0.19		0.75	0.73	0.71
25		0.28	0.25	0.23		0.41	0.35	0.31		0.95	0.94	0.92
50		0.42	0.38	0.35		0.58	0.51	0.49		1.00	1.00	1.00

Note: The nominal size is 5%, and the results are based on 10 000 replications.

Carlo experiment, a fact that is illustrated in Figure 4.1. Even though a large amount of information is implied by a panel with dimensions $T = 100$ and $n = 50$, the panel is not, evidently, large enough.

4.5.2.4 A heterogeneous nonlinear panel with serially correlated errors

The sixth Monte Carlo experiment is conducted to examine the empirical power of the $\mathcal{Z}_3$ test statistic in Proposition 4.4. The DGP is the same as in (4.25) with parameter specifications as in (4.26), but the error process allows for serial correlation and is given by $u_{it} = 0.3u_{i,t-1} + v_{it}$, where $v_{it} \sim nid(0,1)$. The results for the sixth Monte Carlo experiment are presented in Table 4.6.

When the errors are serially correlated, we see that the power for the $\mathcal{Z}_3$ test is reduced compared to the power results for $\mathcal{Z}_0$ with serially uncorrelated errors in Table 4.3. This is partially explained by that in the present case more parameters are estimated. The discrepancy in power between the $\mathcal{Z}_0$ and $\mathcal{Z}_3$ tests is therefore more pronounced for small T, and when increasing T the difference in power between the two tests become modest. Furthermore, ignoring the serial correlation (i.e. the case $p_i = 0$) yields an alarming situation, however not reported here, because the power is then close to zero for all combinations of T and n. Finally note that for $T = 100$, all n, and $p_i \geq 1$, the power is satisfactory.

4.5.2.5 A homogeneous panel with a mixture of nonlinear and nonstationary models

In the seventh and the last Monte Carlo experiment we examine the power when the DGP is a mixture of nonlinear models and unit root processes.

Table 4.7 Empirical power of the $\mathcal{Z}_0$ test in Proposition 4.1 and the IPS test. The DGP is a mixture of LSTAR(1) models and unit root processes.

$n = 25$	(n_1, n_2)											
	$(25, 0)$		$(20, 5)$		$(15, 10)$		$(10, 15)$		$(5, 20)$		$(0, 25)$	
T	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS	$\mathcal{Z}_0$	IPS
25	1.00	0.00	0.99	0.00	0.88	0.00	0.62	0.00	0.20	0.00	0.05	0.05
50	1.00	0.00	1.00	0.00	0.90	0.00	0.64	0.00	0.23	0.00	0.05	0.05
100	1.00	0.01	1.00	0.00	0.96	0.00	0.66	0.00	0.27	0.01	0.05	0.05
250	1.00	0.29	1.00	0.19	1.00	0.16	0.92	0.14	0.59	0.12	0.05	0.05

Note: The nominal size is 5%, and the results are based on 10 000 replications.

The design of the experiments is the following: The nonlinear model is a PLSTAR(1) model with parameters given by (4.27), and the panel unit root processes are given in (4.21). Furthermore, let n_1 and n_2 denote the number of nonlinear models and unit root processes, respectively, such that $n_1 + n_2 = n$ holds. The simulation results are given in Tables 4.7.

From Table 4.7 we conclude that the power of the $\mathcal{Z}_0$ test is decreasing in n_2 towards the nominal size obtained with $n_2 = 25$, cf. also Table 4.1. On the other hand, the power of the $\mathcal{Z}_0$ test is increasing in n_1 and with $n_1 = 25$ the same results as in Table 4.5 are obtained. The power for the $\mathcal{Z}_0$ test is also increasing in T for all fractional combinations of nonlinear and unit root process. Furthermore, it is seen that when a small to medium fraction of the panel is non-stationary, the power of the $\mathcal{Z}_0$ test is barely affected and is close to unity. The power for the IPS test is close to zero because the panel is a mixture of nonlinear and non-stationary processes, and one can not expect the IPS test to gain any power by varying the fraction between these two options.

As a final remark, the results in Table 4.7 also illuminate the important aspect of that the outcome of panel unit root tests should be conservatively dealt with, meaning that rejecting the null hypothesis does not imply that all cross sections are nonlinear. For instance, in the case $n_1 = 5$, $n_2 = 20$, and $T = 250$, the power of the $\mathcal{Z}_0$ test equals 0.59, which means that we might erroneously model the whole panel as nonlinear when in fact only a small fraction of the panel is nonlinear. This also indicates that a careful joint analysis of both the individual and the panel unit root test results is called for to fully assess the stationarity/nonlinearity properties of the panel data, as pointed out in Karlsson and Löthgren (2000).

4.6 Concluding remarks

In this chapter we present a new test for unit roots in a nonlinear dynamic heterogeneous panel. Theirs necessity can be motivated by the fact that canonical panel unit root test, such as the tests in Im, Pesaran, and Shin (2003), do not have satisfactory power when the DGP is an model with shifts in levels. Recent research indicates that many single time series exhibit a nonlinear adjustment path (structural shifts in levels) towards a long-run equilibrium. It seems therefore natural that the cross sections in a panel are modelled in a nonlinear way (or at least a fraction of them) as well.

Our nonlinear dynamic heterogeneous panel is general in the sense that it nests the PLSTAR model in Chapter 3, a panel threshold autoregressive model, as well as the linear autoregressive panel in Im, Pesaran, and Shin (2003). Our panel is characterized by the fact that each cross section is modelled as an LSTAR model and where all parameters are assumed to be cross section specific. The residuals in the PLSTAR model are specified to be independent (a strong assumption) with heterogeneous variance among cross sections, but possibly serially correlated over time.

Our test for a common unit root in a nonlinear panel is based on averaging all individual t-statistics of a unit root for a specific cross section unit. The test is derived under: (i) T and n fixed. (ii) T fixed and letting $n \to \infty$. (iii) First letting $T \to \infty$ and thereafter letting $n \to \infty$ (sequential asymptotics). In the two latter cases it is shown that the limiting distribution of the test is the standard normal distribution.

Monte Carlo studies are performed and it is shown that the size distortions are negligible when the errors are uncorrelated over time. However, when the errors are modelled as an AR(1) process, it is crucial to include sufficiently many lags of the difference of the dependent variable, otherwise the test will be undersized. The power results are very satisfactory and are close to unity for panels being nearly linear or displaying a smooth shifts in levels and dynamics, as long as $n \geq 25$ and $T \geq 50$. In contrast, the IPS test only has power when a linear panel is considered, and otherwise its power is inferior. The improvement in power compared to a univariate test in Chapter 1, is pronounced.

Appendix A

Simulated moments

Table A.1 presents the simulated expected values, variances and third moments for the t_i^m test statistics in (4.8). For comparison, the same moments for the t_i test statistic in (4.7) are calculated as well. The moments are calculated under the null hypothesis in (4.6) assuming that the error term fulfills Assumption 1 with $\sigma_i^2 = 1$ for all i. It is seen that, for all T, $\tilde{S}^2$ is a very accurate estimate of σ_i^2, whereas S^2 is biased downwards. In Table A.1 it is also revealed that the well-known fact of the LS bias in non-stationary regressions, see for instance Abadir (1993), is present in our case as well. Furthermore, we conclude that $\lim_{T\to\infty} \mu(T) = \mu^* = -1$ and $\lim_{T\to\infty} \eta^2(T) = \eta^{2*} = 1.43$, where $\mu(T) = E[t_i^m]$ and $\eta^2(T) = V[t_i^m]$, which confirms that the first two asymptotic moments exist, justifying the use of the Lindberg-Lèvy central limit theorem in Proposition 4.3.

The simulation results for the expected values and variances for t_i^a are reported in Tables A.2 and A.3 respectively. It is pointed out that both the expected value and the variance for t_i^a depends on the factors T, $\mathbf{y}_{i0}$, p_i, $\boldsymbol{\zeta}_i$, and σ_i^2. In order to present these moments, the simulation results are given for various combination of T and p_i when they are both considered as fixed and known. We let $\mathbf{y}_{i0} = \mathbf{0}$ to capture the invariance with respect to σ_i^2. However, the problem with the nuisance parameters $\boldsymbol{\zeta}_i$ under the null hypothesis still remains. The option we choose is to set $\boldsymbol{\zeta}_i = 0$ under the null hypothesis so that data are generated from the model $y_{it} = y_{i,t-1} + u_{it}$ where $u_t \sim nid(0,1)$. On these data we run the regression in (4.18) and calculate the t_i^a statistic in (4.22). This is repeated 1 000 000 times, for each desired sample size T, to generate the expected values and variances for t_i^a. This means that we are approximating e.g. $E[t_i^a(T, p_i, \boldsymbol{\zeta}_i)]$ with $E[t_i^a(T, p_i, \mathbf{0})]$ under the null hypothesis. Without presenting the results we note that these expected

values and variances are rather robust against the values of ζ_i, and that the approximations for the expected values and the variance are reasonable.

For $T \to \infty$ in Tables A.2 and A.3 we see that both the expected value and the variance for t_i^a equal the expected value (i.e. μ^*) and the variance (i.e. η^{2*}) for t_i^m in Table A.1. This is an implication of Corollary 2.7 in Chapter 2.

Table A.1 Simulated moments for t_i and t_i^m.

T	$E[t_i^m]$	$V[t_i^m]$	$E(t_i^m)^3$	$\tilde{S}^2$	$E[t_i]$	$V[t_i]$	$E(t_i)^3$	S^2
10	-0.540	0.898	-1.40	0.997	-0.716	2.086	-4.81	0.649
25	-0.786	1.160	-2.92	0.997	-0.901	1.583	-4.82	0.827
50	-0.889	1.278	-3.84	0.997	-0.950	1.502	-4.95	0.907
100	-0.940	1.361	-4.36	0.998	-0.972	1.466	-4.96	0.951
250	-0.974	1.401	-4.75	0.998	-0.992	1.444	-4.97	0.980
500	-0.985	1.427	-4.87	0.998	-0.995	1.442	-4.98	0.989
1000	-0.991	1.430	-4.97	0.999	-0.997	1.438	-5.00	0.995
∞	-1.000	1.432	-5.03	1.000	-1.000	1.432	-5.03	1.000

Note: The results are based on 1 000 000 replications.

Table A.2 Simulated expected values for t_i^a.

	$E[t_i^a(T, p_i, \boldsymbol{\zeta}_i = \mathbf{0})]$		
p_i	1	2	3
T	ζ_{1i}=0	ζ_{1i}= ζ_{2i} =0	ζ_{1i} = ζ_{2i} = ζ_{3i}=0
25	-0.993	-0.993	-1.052
50	-0.994	-1.000	-1.049
100	-1.000	-1.000	-1.034
250	-1.000	-1.000	-1.011
500	-1.000	-1.000	-1.000
1000	-1.000	-1.000	-1.000
∞	-1.000	-1.000	-1.000

Note: The results are based on 1 000 000 replications.

Table A.3 Simulated variances for t_i^a.

	$V[t_i^a(T, p_i, \boldsymbol{\zeta}_i = \mathbf{0})]$		
p_i	1	2	3
T	$\zeta_{1i} = 0$	$\zeta_{1i} = \zeta_{2i} = 0$	$\zeta_{1i} = \zeta_{2i} = \zeta_{3i} = 0$
25	1.661	1.713	1.762
50	1.553	1.585	1.611
100	1.495	1.513	1.535
250	1.473	1.462	1.484
500	1.451	1.444	1.452
1000	1.441	1.435	1.434
∞	1.432	1.432	1.432

Note: The results are based on 1 000 000 replications.

Bibliography

ABADIR, M. K., (1993): "OLS bias in a nonstationary autoregression," *Econometric Theory,* 9, 81–93.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series," *Journal of the American Statistical Association*, 74, 427–431.

HARRIS, R. D. F., AND E. TZAVALIS (1999): "Inference for unit roots in dynamic panels where the time dimension is fixed," *Journal of Econometrics*, 91, 201–226.

HARVEY, D. I., AND T. C. MILLS (2002): "Unit Roots and Double Smooth Transitions," *Journal of Applied Statistics*, 29, 675–683.

HSIAO, C. (1986): *Analysis of panel data.* Cambridge University Press.

IM, K. S., M. H. PESARAN, AND Y. SHIN (2003): "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115, 53–74.

KARLSSON, S., AND M. LÖTHGREN (2000): "On the power interpretation of panel unit root tests," *Economic Letters*, 66, 249–255.

LANNE, M., H. LÜTKEPOHL, AND P. SAIKKONEN (2003): "Test procedures for unit roots in time series with level shifts at unknown time," *Oxford Bulletin of Economics and Statistics*, 65, 91–115.

LARSSON, R. (1997): "On the asymptotic expectations of some unit root tests in a first order autoregressive process in the presence of a trend," *Annals of the Institute of Statistical Mathematics*, 49, 585–599.

LEVIN, A., C. LIN, AND C. CHU (2002): "Unit root tests in panel data: Asymptotic and finite-sample properties," *Journal of Econometrics*, 108, 1–24.

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

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

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

MADDALA, G. S., AND S. WU (1999): "A Comparative Study of Unit Root Tests With Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, pp. 631–652.

NABEYA, S. (1999): "Asymptotic moments of some unit root test statistics in the null case," *Econometric Theory*, 15, 139–149.

PESERAN, M. H. (2003): "A simple panel unit root test in the presence of cross section dependence," Cambridge University, DAE Working Paper No. 0346.

PHILLIPS, P. C. B., AND H. MOON (1999): "Linear regression limit theory for nonstationary panel data," *Econometrica*, 67, 1057–1111.

PHILLIPS, P. C. B., AND D. SUL (2003): "Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence," Cowles Foundation Discussion Paper No. 1438.

QUAH, D. (1994): "Exploiting cross section variation for unit root inference in dynamic data," *Economic Letters*, 44, 9–19.

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.