FOUR CONTRIBUTIONS TO STATISTICAL INFERENCE IN ECONOMETRICS

Bruno Eklund

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
Fredagen den 5 September 2003
kl. 10.15 i sal Torsten
Handelshögskolan, Sveavägen 65

Four Contributions to Statistical Inference in Econometrics

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics.

EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Organization

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

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

Chairman of the Board: Prof Håkan Lindgren. *Director:* Associate Prof Bo Sellstedt.

Adress

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

FOUR CONTRIBUTIONS TO STATISTICAL INFERENCE IN ECONOMETRICS

Bruno Eklund

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics 2003

ISBN 91-7258-624-9

Keywords: Unit root test, nonlinearity, limit results, heteroskedasticity, density estimation

Printed by:
Elanders Gotab, Stockholm 2003

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

To curiosity

Contents

Acknowledgments

In one way or another, the work with this thesis has benefited from the contributions of many people, and in particular from two persons to whom I forever will be indebted to.

First of all, I would like to express my deepest gratitude to my supervisor Professor Timo Teräsvirta. Above all, for taking me on as a graduate student and giving me the chance to pursue my interests, but also for all support and inspiration during the process of writing this thesis. He has always found the time for discussing or providing me with ideas, and giving suggestions and comments on my work, meanwhile showing great patience with my progress.

Secondly, I would also like to send a special thank to Professor Pentti Saikkonen at the University of Helsinki for pushing me in the right direction when I was starting working on my first two papers, and for providing me with valuable comments on the third paper.

I am also grateful to all members of the Department of Economic Statistic, partly for all comments received in seminars and at the coffee table, but mainly for providing a stimulating and friendly working environment.

Finally, I want to take the opportunity to thank my wife Janka, and my parents, Lars and Birgitta, for always supporting and believing in me.

Part I

Summary of Thesis

Introduction

The last thirty years have been a very productive period in time series analysis and econometrics. With the introduction of the autoregressive moving average (ARMA) model by Box and Jenkins (1970), statisticians received a strong tool which worked well in many situations. The ARMA model is still an important part of the modern statistical analysis, but in contrast with the early years, usually only as an approximation, or a part, of some other model. Nowadays increasing empirical evidence indicates that many features of macroeconomic and financial time series cannot be adequately described and analyzed using linear techniques. It seems to be generally accepted that the economy has nonlinear features and characteristics, and that major economic variables exhibit nonlinear relationships. The analysis and development of different models incorporating various types of nonlinear features has thus become a quickly growing research area during the last two decades. Features such as, structural change, regime shifts and asymmetric behavior of time series can be better described using nonlinear models than the linear ARMA framework.

A common type of nonlinearity, suggested by macroeconomic theory, is regime switching. The switching autoregressive model, described by Tong (1983), has received much attention and numerous studies have applied the model to characterizing nonlinear economic behavior. In the most common case, the threshold model may be viewed as a two-regime system in which each regime specified as a linear model, and where the change between the regimes is assumed to be abrupt. A generalization, allowing for a smooth transition between the regimes, is the smooth transition autoregressive (STAR) model. It also has proved to be a useful tool when modelling nonlinear behavior, see Teräsvirta (1998) for a thorough discussion. An advantage of the STAR model over the threshold model is that the likelihood function of a smooth transition model is continuous everywhere in the parameter space and differentiable. This fact makes it possible to construct tests that rely on standard asymptotic theory. The STAR model has thus received growing attention during the last

few years.

The general theme of this thesis is nonlinear statistical inference. The first two chapters combine nonstationarity and STAR: the idea is to develop unit root tests against a stationary STAR alternative. There are two applications: the first one is to the monthly seasonally adjusted US unemployment rate, and the second has to do with testing the purchasing power parity hypothesis with a large number of real exchange rates. The third chapter considers the problem of testing constancy of a covariance matrix of a vector model, such as a vector autoregressive model. The VAR modellers rarely test their models, but the test developed in this thesis offers an easy way of checking whether or not the model is correctly specified. Finally, the last chapter considers a way of estimating confidence regions nonparametrically in situations where a simple linear grid is not a feasible solution. The application consists of estimating the confidence region for a pair of kurtosis and first-order autocorrelation of squared observations of a first-order GARCH model. A more detailed overview of the chapters follows next.

Summary and the main results of the papers

Testing the unit root against nonlinear models

There are situations in which the standard Dickey-Fuller tests do not function as well as one might desire. Difficulties in detecting structural change, shifts in mean or growth rate, or nonlinear behavior have been noticed in several studies. Nelson and Plosser (1982) showed for a number of financial and macroeconomical time series that the Dickey-Fuller tests are unable to reject the unit root hypothesis. This paper started lively research on unit root tests such that the alternative to the unit root was assumed not to be a stationary autoregressive model. Pippenger and Goering (1993) argued that examining long-run economic relationships using the unit root tests is questionable in the presence of transaction costs or hysteresis thresholds. Perron (1989) argued that low power against structural breaks in level and growth rate can result in overstating the evidence in favor of unit roots. However, the standard Dickey-Fuller tests are also affected by size distortions in a number of situations. Size distortions, in the form of too frequent rejections of the null, have been observed when there is a single structural break in trend or variance under the null hypothesis, see Leybourne, Mills and Newbold (1998) and Hamori and Tokihisa (1997). Similar size distortion, when the true null model contains Markov regime switching in trend growth rate, as demonstrated by Nelson, Piger and Zivot (2001). They also showed power loss testing the unit root hypothesis was tested against a true alternative process with a Markov-switching trend.

The emphasis in these studies, and in a major part of the early literature, has been on the linear model which today is increasingly viewed to be somewhat inadequate. The emphasis of linear models has led to downplaying potential nonlinear characteristics of many time series. The increasing empir-

ical evidence on nonlinear relationships and features in economical time series during the last few years has resulted in efforts to incorporate nonlinear models and techniques into the existing econometric framework. The literature on testing the unit root hypothesis against nonlinear models, or vice versa, has recently grown rapidly in this direction.

One of the more recent studies in this area is Caner and Hansen (2001) who analyzed and provided tests of the unit root hypothesis against the threshold autoregressive (TAR) model. The authors also proposed a bootstrap method to approximate the limit distribution of the test under the null, and showed that the unit root hypothesis can be rejected for the U.S. unemployment rate in favor of the nonlinear TAR model. Also considering the TAR model as an alternative to the unit root, Enders and Granger (1998) found that movements toward long-run equilibrium relationship of an interest rate are best described as an asymmetric process. Modifying the test in Enders and Granger (1998), Berben and van Dijk (1999) found evidence of asymmetric adjustments towards long-run equilibrium for a number of forward premium time series. Kapetanios and Shin (2000) developed and analyzed the unit root test with the self-exciting threshold autoregressive (SETAR) model as the alternative process. Designed to take into account the threshold nature under the alternative, they reported some gain in power compared to the Dickey-Fuller test. Other studies have integrated the unit root test with nonlinear models with a smooth transition between the regimes. Kapetanios, Shin and Snell (2003) proposed a test of the joint unit root and linearity hypothesis against a very simple exponential smooth transition autoregressive (ESTAR) model that only allows a regime shift in the slope parameter. They were able to reject the unit root for a number of the real interest rates in favor of the ESTAR model. Bec, Salem and Carrasco (2002) tested the unit root hypothesis against a logistic STAR model with three regimes and were able to reject the unit root for eleven out of 28 real exchange rates considered.

Chapter 1. Testing the unit root hypothesis against the logistic smooth transition autoregressive model

The first paper contains statistical theory for testing the unit root hypothesis against the smooth transition autoregressive (STAR) model. New limit distribution results are provided, together with two F type test statistics for the joint unit root and linearity hypothesis against a specific nonlinear alternative. Nonlinearity is defined through the smooth transition autoregressive (STAR) model. The joint hypothesis of a unit root and linearity allows one to distinguish between random walk processes, with or without drift, and sta-

tionary nonlinear processes of the smooth transition autoregressive type. Due to occasional size distortion in small samples, a simple bootstrap method is proposed for estimating the p-values of the tests. Power simulations show that the two tests, F_{nd} and F_d, have at least the same or higher power than the corresponding Dickey-Fuller tests. For illustration, the tests are applied to the seasonally adjusted U.S. monthly unemployment rate. The linear unit root hypothesis is strongly rejected, which suggest considerable evidence that the unemployment series is better described by a stationary smooth transition autoregressive process than a random walk. The test result is of interest when the possible presence of hysteresis in the U.S. unemployment is considered.

Chapter 2. A nonlinear alternative to the unit root hypothesis

This second paper extends the work in chapter 1. Two tests are constructed for the joint linearity and unit root hypothesis against the second-order logistic STAR model. Compared to the work in chapter 1, the alternative STAR model in this paper allows for regime shifts in intercept, growth rate and in level. Some new limit results are presented and two F type tests are proposed. Small sample simulations show some size distortions, and for this reason the bootstrap method from chapter 1 is applied also here. Power simulations show some gain in power, compared to the standard Augmented Dickey-Fuller tests. Finally, the two proposed F type tests are applied to a number of real exchange rates in an attempt to test the purchasing power parity (PPP) hypothesis. For 44 out of 120 exchange rates considered, the linear unit root is rejected in favor of the stationary nonlinear STAR model. These results of the tests supports the PPP hypothesis, even though the test results, at best, can be considered as a complement to earlier studies.

Chapter 3. Testing the constancy of the error covariance matrix in vector models

Estimating the parameters of an econometric model is necessary for any use of the model, be it forecasting or policy evaluation. Finding out thereafter whether or not the model appears to satisfy the assumptions under which it was estimated should be an integral part of a normal modelling exercise. An evaluation through misspecification tests has long been standard practice in univariate time series analysis, see Box and Jenkins (1970). It has been advocated for multivariate regression models as well, and a number of misspecification tests exist for linear and nonlinear single-equation models.

Misspecification tests also exist for estimated vector models such as the linear vector autoregressive (VAR) model, although their use in the VAR context seems much less widespread than in the single-equation case. The available tests are generally intended for testing the specification of the conditional mean. It is argued, however, that sometimes misspecification of the conditional mean shows in the error variance, making it time-varying. Besides, the error variance of a model may be nonconstant even when the conditional mean is correctly specified. It would therefore be desirable to have statistical tools for investigating possible parameter nonconstancy in vector models.

Many approaches to testing the constancy of the error variance in single-equation models exist in the literature, whereas the same is not true for constancy of the error covariance matrix in a multivariate system. A special case of the test discussed in this paper may be viewed as a multivariate generalization of the heteroskedasticity test of White (1980), and another special case generalizes the test against autoregressive conditional heteroskedasticity of Engle (1982). Yet another variant of the test generalizes the constant variance test of Medeiros and Veiga (2003) who assumed that under the alternative, the error variance changes smoothly over time. An important assumption in the present paper is that while the error variances change over time, the correlations between them remain constant. This restriction has the advantage that it considerably decreases the dimension of the null hypothesis compared to the case where both variances and covariances can fluctuate freely under the alternative hypothesis. This paper contains a Lagrange multiplier (LM) test of the hypothesis that the covariance matrix of a multivariate time series model is constant over time. It is further assumed that under the alternative, the error variances are time-varying whereas the correlations remain constant over time. Under the parameterized alternative hypothesis the variance may change continuously as a function of time or some observable stochastic variables. A rejection of the null hypothesis against a particular parametric alternative does not necessarily mean that the true model is the specified one with errors following this alternative. It could as well be an indication on a misspecified conditional mean or caused by a structural change, and the test presented here should therefore be best viewed as a useful diagnostic test.

Monte Carlo simulations show that the LM test has good size properties already in small samples. No major size distortion is detected in any of the experiments considered. The power simulations show that the test has very good power against a correctly specified alternative, but low or only up to moderate power in cases for a misspecified alternative hypothesis. Simulations have also been carried out in the case where the assumption of constant correla-

tions is violated. The result show that the test is not affected by time-varying correlations in any major extent, except for a very particular case where the correlation increase monotonically over time and the variance is assumed to change smoothly with time under the alternative.

Chapter 4. Estimating confidence regions over bounded domains

In classical statistical analysis, the inference rely primarily on an estimated parametric model. Hypothesis testing can then easily be performed, and confidence regions and forecasts obtained using the data and the model. However, in situations where the parametric form of the model is unknown, other methods have to be applied to provide the appropriate tools for statistical inference when analyzing a data set. Suppose we want to estimate a joint density of two random variables but do not know the form of the density. In such situations nonparametric density estimation by kernel techniques is a standard statistical tool. This paper has to do with a special situation where the observations used for estimating the density are obtained from an unknown member of a known family of models. Information about the properties of the model may then be used in estimating the density function of interest. It may be a joint density of some random variables that in turn are functions of observations generated by a model that belongs to a known family of models.

The example motivating this paper is one in which the problem is to estimate confidence regions for the combination of kurtosis and first-order autocorrelation of squared observations of a generalized autoregressive heteroskedasticity (GARCH) process. Such confidence regions can be obtained nonparametrically by first estimating the joint density of the kurtosis and autocorrelation estimators from a random sample. The interest in this problem originates from considerations in Teräsvirta (1996) who discussed the ability of some simple volatility models to reproduce so-called stylized facts evident in financial return series. As it turns out, the relationship between the two estimators is nonlinear, which in turn implies that their joint density has such a form that standard density estimation techniques based on a linear grid do not work properly. A standard technique in such cases would be to transform the data to simplify the estimation of the density. However, in many situations, as in the present case, no one-to-one transformation of the considered data set is known to exist, or if it existed, the existence would be difficult to verify. A solution would be to apply some method which takes into account the nonlinear dependence between the estimators. The topic of this paper is

to propose such a method. The main idea is to, instead of transforming the data to be able to use the linear grid, transform the grid itself to better represent the data. By taking into account parameter restrictions of an underlying model, the GARCH model in this case, a nonlinear grid can be constructed over which the density function can be obtained by simple kernel estimation. As an illustration the method is applied to the kurtosis/first-order autocorrelation of squared observations of three GARCH(1, 1) models. For each model, the resulting confidence region covers a reasonable area of the definition space and is well aligned with the data set.

Bibliography

[1] Bec, F., Salem, M. B., and Carrasco, M., (2002), "The purchasing power parity puzzle: Evidence from a LSTAR specification", Working Paper, Université de Paris I.

[2] Berben, R.-P., and van Dijk, D., (1999), "Unit root tests and asymmetric adjustment: A reassessment", Working paper, Erasmus University Rotterdam.

[3] Box, G. E. P., and Jenkins, G. M., (1970), *Time series analysis: forecasting and control*, San Francisco, Holden Day.

[4] Caner, M., and Hansen, B. E., (2001), "Threshold autoregression with a unit root", *Econometrica*, 69, 1555-1596.

[5] Enders, W., and Granger, C. W. J., (1998), "Unit-root tests and asymmetric adjustment with an example using the term structure of interest rate", *Journal of Business and Economic Statistics*, 16, 304-311.

[6] Engle, R. F., (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", Econometrica, 50, 987-1007.

[7] Hamori, S., and Tokihisa, A., (1997), "Testing for a unit root in the presence of a variance shift", *Economics Letters*, 57, 245-253.

[8] Kapetanios, G., and Shin, Y. (2000), "Testing for a linear unit root against nonlinear threshold stationarity", Discussion paper, http://www.econ.ed.ac.uk/Academic_00Discuss.htm, University of Edingburgh.

[9] Kapetanios, G., Shin, Y., and Snell, A., (2003), "Testing for a unit root in the nonlinear STAR framework", *Journal of Econometrics*, 112, 359-379.

[10] Leybourne, S. J., Mills, T. C., and Newbold, P., (1998), "Spurious rejections by Dickey-Fuller tests in the presence of a break under the null", *Journal of Econometrics*, 87, 191-203.

[11] Medeiros, M. C., and Veiga, A., (2003), "Diagnostic checking in a flexible nonlinear time series model", *Journal of Time Series Analysis*, (forthcoming).

[12] Nelson, C. R., Piger, J., and Zivot, E., (2001), "Markov regime switching and unit root tests", *Journal of Business and Economic Statistics*, 19, 404-415.

[13] Nelson, C. R., and Plosser, C. I., (1982), "Trends and random walks in macroeconomic time series", *Journal of Monetary Economics*, 10, 139-162.

[14] Perron, P., (1989), "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 57, 1361-1401.

[15] Pippenger, M. K., and Goering, G. E., (1993), "A note on the empirical power of unit root tests under threshold processes", *Oxford Bulletin of Economics and Statistics*, 55, 473-481.

[16] Teräsvirta, T., (1996), "Two stylized facts and the GARCH(1, 1) model", Stockholm School of Economics, *Working Paper Series in Econometrics and Finance*, No. 96.

[17] Teräsvirta, T., (1998), "Modeling economic relationships with smooth transition regressions", in A. Ullah and D. E. A. Giles (eds.), *Handbook of applied economic statistics*, Dekker, New York, 507-552.

[18] Tong, H., (1983), *Threshold models in non-linear time series analysis*, New York: Springer-Verlag.

[19] White, H., (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 48, 817-838.

Part II

The chapters

Chapter 1

Testing the unit root hypothesis against the logistic smooth transition autoregressive model

1.1 Introduction

The research and development of unit root tests have been considerable for the past twenty years. A main motivation has been to analyze and explain the allegedly poor properties of the standard unit root tests and to propose alternative tests, often in an attempt to increase the applicability and power of the tests. As Nelson and Plosser (1982) showed, standard Dickey-Fuller tests are unable to reject the hypothesis of a unit root in several macroeconomic and financial time series. The poor size and power properties of the standard tests when the time series contain structural changes, shifts in mean or growth rate, or nonlinear behavior, have been noticed in several studies. Pippenger and Goering (1993) showed how the power of the standard Dickey-Fuller tests falls considerably when the true alternative is a threshold autoregressive (TAR) model. They pointed out that in the presence of transaction costs or hysteresis thresholds the usefulness of standard unit root tests in examining long-run economic relationships is suspect. Diebold and Rudebusch (1990) showed similar loss in power when the true alternative is a fractionally integrated process. Perron (1989) argued that the low power against structural breaks in level and growth rate can result in overstating the evidence in favor of unit roots. The converse problem does, however, also exist, that standard unit root tests reject too often when there is a single structural break in trend or variance under the null hypothesis, as demonstrated by Leybourne, Mills and Newbold (1998), and Hamori and Tokihisa (1997). Nelson, Piger and Zivot (2001) showed similar results of size distortion when the true null model contains Markov regime switching in trend growth rate. They also showed low power testing the unit root hypothesis against an alternative process with Markov-switching trend.

Since the main focus of these studies was on analyzing the linear model, any possible nonlinear properties or features of the time series were ignored. On the other hand, there exists empirical evidence indicating that many features of macroeconomic and financial time series cannot be adequately described and analyzed using linear techniques. As a result, nonlinear models have become an active area of research in econometrics. Among other things, interest has been devoted to the problem of testing the joint hypothesis of linearity and unit root of a time series against specific nonlinear and stationary alternatives. The literature in this area has been growing rapidly.

An example of a recent study of this kind is Enders and Granger (1998). The authors analyzed and provided a test of the unit root hypothesis against the threshold autoregressive (TAR) model. They found that movements toward long-run equilibrium relationship of an interest rate are best estimated as

an asymmetric process. Berben and van Dijk (1999), who applied a modified version of the Enders and Granger (1998) test, found asymmetric adjustments in several forward premium series. Caner and Hansen (2001), who also considered the TAR model as an alternative to the unit root hypothesis, proposed a bootstrap procedure to approximate the sampling distribution of the test statistic under the null. They reported strong evidence that U.S. male unemployment is better described by a stationary TAR process than a unit root process. Further examples are Kapetanios and Shin (2000), who developed and analyzed a test against the self-exciting threshold autoregressive (SETAR) model. Their test was more powerful than the Dickey-Fuller test that ignores the threshold nature under the alternative. Kapetanios, Shin and Snell (2003) considered a simple exponential smooth transition autoregressive model, only allowing for a regime shift in the slope parameter, as the alternative to the joint linearity and unit root hypothesis. As an illustration they provided an application to real interest rates, and rejected the null hypothesis for several interest rates considered, whereas the standard Augmented Dickey-Fuller tests failed to do that.

This paper will consider testing the joint linearity and unit root hypothesis against a smooth transition autoregressive (STAR) model. Standard STAR models has two extreme regimes, and the transition between them is smooth; see Teräsvirta (1994) for more discussion. Furthermore, the two-regime TAR model is included in the STAR model as a special case.

The paper will be organized as follows. Section 2 contains the model specification. Asymptotic results, limiting distributions for the two resulting test statistics and critical values are provided in section 3. Section 4 describes the bootstrap method to estimate p-values, and reports results of Monte Carlo simulations of the small sample properties, size and power, of the proposed tests. Section 5 contains an empirical application, and Section 6 concludes. The appendix includes proofs of the two theorems.

A few words on the notation. All limits are taken as $T \longrightarrow \infty$, and weak convergence is denoted as $\Rightarrow$.

1.2 Model, null hypothesis and auxiliary regression

Consider the following univariate smooth transition autoregressive (STAR) model

$$\Delta y_t = \theta_0 + \theta_1 \Delta y_{t-1} + \psi y_{t-1} + (\varphi_0 + \varphi_1 \Delta y_{t-1}) F(\gamma, c, y_{t-1}) + \varepsilon_t, \qquad (1.1)$$

where $t = 1, \ldots, T$, and the dependent variable, y_t, is included in the model both as differences, Δy_t, and levels, y_t, but in the function $F(\cdot)$ only in levels. The differences, Δy_t, and errors, ε_t, are assumed to be stationary, and the errors satisfy $E\varepsilon_t = 0$. Furthermore, the transition function $F(\cdot)$ is a bounded continuous function, $F(\cdot) \in [-\frac{1}{2}, \frac{1}{2}]$. This allows the dynamic behavior of the model to change between two regimes corresponding to the cases when $F(\cdot) = -\frac{1}{2}$ and $F(\cdot) = \frac{1}{2}$, smoothly with the transition variable y_{t-1}. A number of different possibilities exist for the choice of the function $F(\cdot)$, see Granger and Teräsvirta (1993) or Teräsvirta (1998) for a presentation and discussion of the most common functional forms. This paper will focus on the logistic function

$$F(\gamma, c, y_{t-1}) = (1 + \exp(-\gamma(y_{t-1} - c)))^{-1} - \frac{1}{2}, \tag{1.2}$$

where the parameter c is the transition midpoint parameter, and γ the speed of transition from one extreme regime to the other. Note that as $\gamma \longrightarrow \infty$ the function $F(\cdot)$ approaches a step function, so the model ultimately becomes a threshold autoregressive (TAR) model, see Tong (1983). On the other hand, when $\gamma = 0$ the function $F(\cdot)$ is constant for all values of y_{t-1}, implying that model (1.1) is linear. The linearity hypothesis can thus be defined as $\gamma = 0$.

Testing $H_0 : \gamma = 0$ in (1.1) and (1.2) is not straightforward, however. The reason for this is the identification problem as (1.1) is only identified for $\gamma > 0$ but not for $\gamma = 0$, see Luukkonen, Saikkonen and Teräsvirta (1988), Teräsvirta (1994a,b), Lin and Teräsvirta (1994) for details. Following the idea in Luukkonen et al. (1988), this problem can be circumvented by a first-order Taylor approximation around $\gamma = 0$ in $F(\gamma, c, y_{t-1})$. This results in the following approximate model:

$$\begin{aligned}
\Delta y_t &= \theta_0 + \theta_1 \Delta y_{t-1} + \psi y_{t-1} + (\varphi_0 + \varphi_1 \Delta y_{t-1}) \frac{\gamma}{4}(y_{t-1} - c) + \varepsilon_t^* \\
&= \left(\theta_1 - \frac{\varphi_1 \gamma c}{4}\right) \Delta y_{t-1} + \frac{\varphi_1 \gamma}{4} y_{t-1} \Delta y_{t-1} + \left(\theta_0 - \frac{\varphi_0 \gamma c}{4}\right) \\
&\quad + \left(\psi + \frac{\varphi_0 \gamma}{4}\right) y_{t-1} + \varepsilon_t^* \\
&= \delta \Delta y_{t-1} + \phi y_{t-1} \Delta y_{t-1} + \alpha + \zeta y_{t-1} + \varepsilon_t^*,
\end{aligned} \tag{1.3}$$

where $\varepsilon_t^* = \varepsilon_t + R_1(\gamma, y_{t-1})(\varphi_0 + \varphi_1 \Delta y_{t-1})$, R_1 being the remainder. In equation (1.3), the linearity hypothesis now corresponds to $\phi = 0$. Also note that under the null hypothesis $\varepsilon_t^* = \varepsilon_t$ since the remainder $R_1 \equiv 0$. Moving y_{t-1} to the right hand side results in the following model:

$$y_t = \delta \Delta y_{t-1} + \phi y_{t-1} \Delta y_{t-1} + \alpha + \rho y_{t-1} + \varepsilon_t^*, \tag{1.4}$$

where $\rho = \zeta + 1$. When the additional term $y_{t-1}\Delta y_{t-1}$ is excluded, this auxiliary autoregression is the model used in the Augmented Dickey-Fuller (ADF) test with a constant and one lag of Δy_t as regressors. As a consequence, if $\varphi_1 = 0$ in the STAR model (1.1), the resulting auxiliary and ADF models are indistinguishable from each other, even for $\varphi_0 \neq 0$. No additional power, compared to the ADF tests, can therefore be expected of the two tests to be proposed below if there is a regime shift only in the intercept. A remedy to this problem would be to base the tests on a third-order Taylor approximation to (1.2), as in Teräsvirta (1994a,b). This case will not, however, be considered any further here.

In Kapetanios et al. (2003), the corresponding auxiliary autoregression had the form

$$\Delta y_t = \delta y_{t-1}^3 + \varepsilon_t, \tag{1.5}$$

indicating that they have a random walk without drift or time trend under their null hypothesis, $\delta = 0$. This was also implied by their alternative model, but by using de-meaned and de-trended variables in a two-step procedure they were able to allow for a random walk with drift and a random walk with time-trend under the null.

Since model (1.4) nests the ADF model, this specification makes it possible to set up a joint test of the unit root and linearity hypotheses, allowing y_t to follow a stationary nonlinear process under the alternative. A joint test of the unit root and linearity hypotheses against nonlinearity amounts to testing the hypothesis $H_{01} : \phi = \alpha = 0$, $\rho = 1$ in (1.4). It is easily seen that y_t in fact has a unit root under this null hypothesis, since model (1.4) then equals:

$$y_t = \delta\Delta y_{t-1} + y_{t-1} + \varepsilon_t. \tag{1.6}$$

Equation (1.6) can also be written as

$$\Delta y_t = \delta\Delta y_{t-1} + \varepsilon_t, \tag{1.7}$$

or, equivalently, as an infinite-order moving average model

$$\Delta y_t = (1 - \delta L)^{-1}\varepsilon_t = \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} = \omega(L)\,\varepsilon_t = u_t, \tag{1.8}$$

where L is the lag operator, that is, $Ly_t = y_{t-1}$. Thus, under H_{01}, $\{y_t\}$ is a random walk without drift. Under the maintained hypothesis of a unit root under the null, the hypothesis $H_{02} : \phi = 0$, $\rho = 1$ corresponds to the case of a random walk with drift.

From equation (1.3) it is seen that testing the hypothesis $\alpha = 0$ implies a test of $\theta_0 = 0$ in the STAR model (1.1) under the original linearity condition $\gamma = 0$. Also, by the same token, restriction $\rho = 1$ implies a test of $\psi = 0$. Thus, when the null hypothesis H_{01} is rejected, parameters θ_0 and ψ must be included in the alternative model.

Since the Δy_t process is assumed to be stationary, equation (1.7) implies that the parameter $|\delta| < 1$. Note that if $\delta = 1$, Δy_t has a unit root so that y_t is an $I(2)$ process. On the other hand, if $\delta = -1$, y_t has a negative unit root. From this follows that problems can arise in practice if δ is close to -1 or 1. This problem will be analyzed and discussed in section 4.

1.3 Limit results and the asymptotic tests

Let $b_T = \left(\widehat{\delta}, \widehat{\phi}, \widehat{\alpha}, \widehat{\rho}\right)'$ be the ordinary least squares estimator of $\beta = (\delta, \phi, \alpha, \rho)'$ in (1.4), so that

$$b_T - \beta = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t, \tag{1.9}$$

where $x_t = (\Delta y_{t-1}, y_{t-1}\Delta y_{t-1}, 1, y_{t-1})'$. The convergence rates and limit distributions for most of the elements in the matrix and vector in equation (1.9) are known from previous studies. However, probability limits to some of the terms have not been considered before and are given in Theorem 1. First, the following assumption, employed by Hansen (1992), is assumed to be satisfied.

Assumption 1 *For some $q > \eta > 2$, $\{v_t\}$ is a zero mean, strong mixing sequence with mixing coefficients of size $-q\eta/(q-\eta)$, and $\sup_{t\geq 1} \|v_t\| = C < \infty$. In addition,*

$$T^{-1}E\left(V_T V_T'\right) \longrightarrow \Omega < \infty \text{ as } T \longrightarrow \infty, \text{ where } V_T = \sum_{t=1}^{T} v_t. \tag{1.10}$$

This assumption allows for a wide variety of different mixing processes, and in particular processes with weakly dependent heterogeneous observations that are common in econometric applications. The following result can now be stated:

Theorem 1 *Assume that $\{u_t\}$ in equation (1.8) satisfies Assumption 1, and let $\{\varepsilon_t\}$ be an i.i.d. sequence with mean zero, variance σ^2, and a finite fourth*

moment. Define

$$\gamma_j = E(u_t u_{t-j}) = \sigma^2 \sum_{s=0}^{\infty} \omega_s \omega_{s+j} \quad , j = 0, 1, \ldots$$
$$\lambda = \sigma \sum_{j=0}^{\infty} \omega_j = \sigma\omega(1)$$
$$\xi_t = \sum_{i=0}^{t} u_i \qquad , t = 1, 2, \ldots, T,$$

with $\xi_0 = 0$. *Then the following expressions converge jointly:*

$$(a) \quad T^{-3/2} \sum_{t=1}^{T} \xi_t u_t^2 \Rightarrow \gamma_0 \lambda \int_0^1 W(r)\, dr$$
$$(b) \quad T^{-3/2} \sum_{t=1}^{T} \xi_t u_t^3 \Rightarrow E\left(u_t^3\right) \lambda \int_0^1 W(r)\, dr$$
$$(c) \quad T^{-2} \sum_{t=1}^{T} \xi_t^2 u_t^2 \Rightarrow \gamma_0 \lambda^2 \int_0^1 W^2(r)\, dr$$
$$(d) \quad T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_{t-1} \varepsilon_t \Rightarrow \sigma\sqrt{\gamma_0}\lambda \int_0^1 W(r)\, dB(r)$$
$$(e) \quad T^{-2} \sum_{t=1}^{T} \xi_t^2 u_t = o_p(1)$$

where $W(r)$ *and* $B(r)$ *are two independent standard Brownian motions defined for* $r \in [0, 1]$.

Proof. See the Appendix.

Observing the rates of convergence in Theorem 1 and other known limit results allows one to define the scaling matrix

$$\Upsilon_T = diag\left(T^{1/2}, \ T, \ T^{1/2}, \ T \right). \tag{1.11}$$

Then, pre-multiplying $b_T - \beta$ in equation (1.9) by Υ_T, finite limits to the ordinary least squares estimates are given by

$$\Upsilon_T (b_T - \beta) = \left\{ \Upsilon_T^{-1} \left(\sum_{t=1}^{T} x_t x_t' \right) \Upsilon_T^{-1} \right\}^{-1} \left\{ \Upsilon_T^{-1} \left(\sum_{t=1}^{T} x_t \varepsilon_t \right) \right\}. \tag{1.12}$$

The null hypothesis, $H_{01} : \phi = \alpha = 0, \rho = 1$, has the alternative representation $H_{01} : R\beta = r$, where $R = \left[\ 0_3 \quad I_3 \ \right]$, $0_3 = (0, 0, 0)'$, $\beta = (\delta, \phi, \alpha, \rho)'$ as before,

and $r = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}'$. The F test statistic is then defined in the usual way as

$$F = (b_T - \beta)' (R\Upsilon_T)' \left\{ s_T^2 R\Upsilon_T \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \Upsilon_T R' \right\}^{-1} R\Upsilon_T (b_T - \beta) / k, \tag{1.13}$$

where

$$s_T^2 = \frac{1}{T-4} \sum_{t=1}^{T} \left(y_t - \widehat{\delta} \Delta y_{t-1} - \widehat{\phi} y_{t-1} \Delta y_{t-1} - \widehat{\alpha} - \widehat{\rho} y_{t-1} \right)^2 \tag{1.14}$$

is a consistent estimator of σ^2, and k equals the number of restrictions. In the present case, $k = 3$. The corresponding statistic for testing the hypothesis H_{02} that allows a drift term is obtained by setting

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, r = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1.15}$$

and $k = 2$. The test statistics will be called F_{nd} and F_d, where *nd* and *d* stand for '*no drift*' and '*drift*', respectively. Obviously, they do not have standard asymptotic distributions as is the case in testing linearity in stationary STAR processes. Their asymptotic distribution theory under the two null hypotheses is given in the following theorem:

Theorem 2 *Let Assumption 1 and the results of Theorem 1 hold. Then the test statistics, F_{nd} and F_d, will have the following asymptotic distributions under the two null hypotheses;*

(*i*) *Under* H_{01} : $\phi = \alpha = 0, \rho = 1$,

$$F_{nd} \Rightarrow \frac{1}{3} W^2(1) + \frac{1}{3} \frac{\left(W(1) \int_0^1 W(r)\, dr - \int_0^1 W(r)\, dB(r) \right)^2}{\int_0^1 W^2(r)\, dr - \left(\int_0^1 W(r)\, dr \right)^2} + \frac{1}{3} \frac{\left(W(1) \int_0^1 W(r)\, dr - \frac{1}{2} \{ W^2(1) - 1 \} \right)^2}{\int_0^1 W^2(r)\, dr - \left(\int_0^1 W(r)\, dr \right)^2}. \tag{1.16}$$

$$(ii) \qquad \textit{Under } H_{02} \quad : \quad \phi = 0,\ \rho = 1,$$

$$F_d \Rightarrow \frac{1}{2}\frac{\left(W(1)\int_0^1 W(r)\,dr - \int_0^1 W(r)\,dB(r)\right)^2}{\int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2} + \frac{1}{2}\frac{\left(W(1)\int_0^1 W(r)\,dr - \frac{1}{2}\left\{W^2(1)-1\right\}\right)^2}{\int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2}. \tag{1.17}$$

Proof. See the appendix.

Asymptotic critical values for the F_{nd} and F_d statistics have been generated by simulation. This has been done by estimating the two F-type test statistics from observations generated from the null model (1.6) with $\delta = 0$ at sample sizes $T = 25, 50, 100, 250, 500$ with 1000000 replications. These values can be found in Table 1.1 where quantiles for the asymptotic distributions of the tests, (1.16) and (1.17), are also included. They have been estimated using the same number of replications as before with $T = 10000$ observations.

Table 1.1 Critical values for the test statistics F_{nd} and F_d, $\delta = 0$.

	F_{nd}					F_d				
T	0.10	0.05	0.025	0.01	0.001	0.10	0.05	0.025	0.01	0.001
25	3.24	4.04	4.87	6.02	9.34	4.23	5.38	6.58	8.25	12.99
50	3.09	3.76	4.43	5.33	7.72	4.08	5.06	6.05	7.36	10.81
100	3.04	3.66	4.27	5.07	7.05	4.04	4.96	5.85	7.03	10.01
250	3.02	3.62	4.20	4.95	6.79	4.03	4.92	5.78	6.90	9.63
500	3.02	3.61	4.18	4.91	6.72	4.03	4.92	5.77	6.86	9.55
∞	3.00	3.58	4.14	4.86	6.62	4.03	4.90	5.74	6.83	9.43

Note that even if the limit distributions do not depend on any nuisance parameters, the critical values for small sample sizes do. Under the null hypothesis, y_t is a function of δ, as is seen from equation (1.6). Thus, under perfect conditions, with δ known, critical values can be easily estimated. This is of course not normally the case in practice. As noted above, special attention is needed for values of δ close to -1 or 1. The time series Δy_t is then close to having a unit root or becoming nonstationary. In these situations the

test may reject the null hypothesis too often. This property of the tests will be investigated in detail in the next section.

1.4 Small sample properties of the tests

In this section both the size and the power properties of these statistics are examined. For comparison, this is also done for the corresponding ADF tests, here called ADF_{nd} and ADF_d. Furthermore, a bootstrap method for adjusting the size of the tests in small samples is proposed.

1.4.1 Size simulations

To investigate the size properties the rejection probabilities of the true null hypothesis are computed, using Monte Carlo simulations and corresponding critical values from Table 1.1, for sample sizes $T = 25, 50, 100, 250, 500, 1000$. The nominal size considered for each test is 5%, and the number of replications equals 1000000. The data are generated from the null model (1.6) assuming $\{\varepsilon_t\} \sim nid(0, 1)$. Since the empirical size in small samples is affected by δ, it has been computed for a number of different values of δ ranging from minus one to one. Figures 1.1 to 1.4 display the estimated sizes of the F_{nd}, ADF_{nd}, F_d and ADF_d test statistics as functions of δ for the different sample sizes. In the two ADF tests, the regression equation contains the first lag of Δy_t.

The test statistics whose empirical sizes are shown in Figures 1.1 to 1.4 all share the characteristic that their size properties become poor for values of δ close to one. Furthermore, the test statistics F_{nd} and F_d are oversized for values of δ near -1 as well. As noted above, $\{y_t\}$ is $I(2)$ for $\delta = 1$, which manifests itself in high rejection frequencies of the true null hypothesis when δ is near unity. For $|\delta| = 1$ the stationarity assumption of Δy_t is violated, which, as can be expected, results in more frequent rejections of the null hypothesis than the asymptotic theory prescribes. In the figures, the smaller the sample size, the larger the deviation from the nominal 5% size level. The size tends towards the nominal 5% when the sample size is increased. Another notable fact is that the deviations from the nominal size for the two tests allowing for a drift term, F_d and ADF_d, are not as large as the ones for the F_{nd} and ADF_{nd}. As the critical values in Table 1.1 are estimated for $\delta = 0$, calculating new critical values or using bootstrapped p-values in the tests is recommended whenever δ is suspected to be close to -1 or 1 and the sample size is small.

Figure 1.1 Size of the F_{nd} statistic for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.

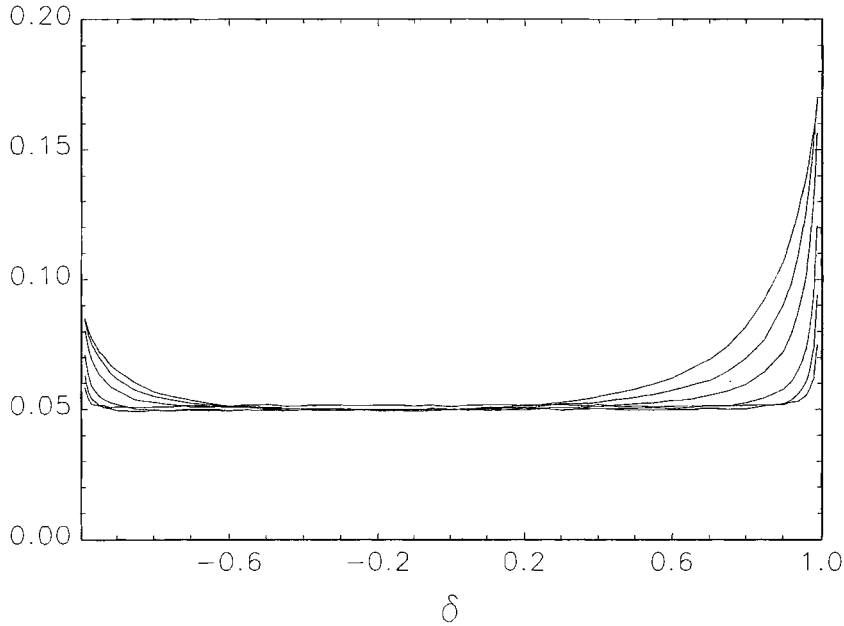
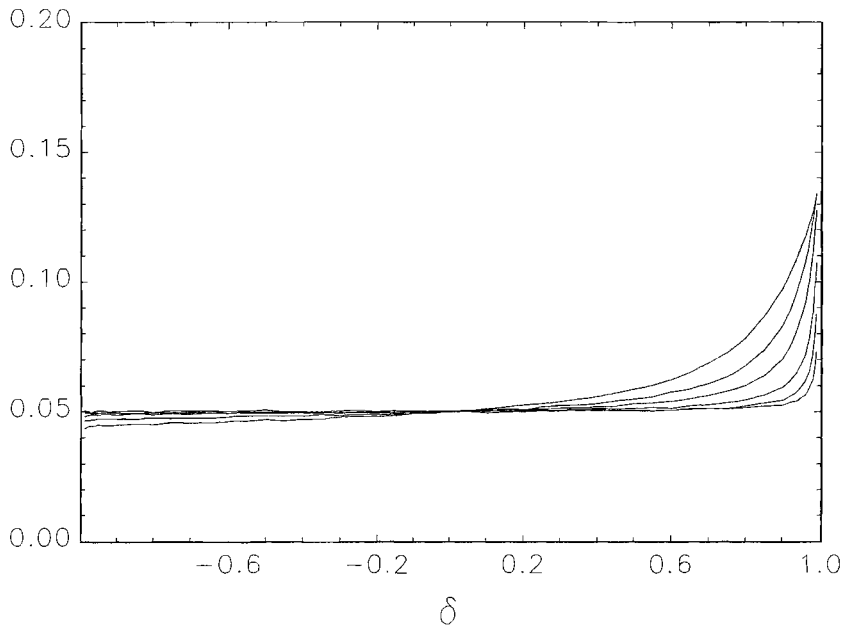


Figure 1.2 Size of the ADF_{nd} statistic for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.



The empirical size of F_{nd} and F_d have also been estimated for two cases of non-normal errors. Considering errors drawn from $t(6)$ and $\chi^2(1) - 1$ the Monte Carlo simulation setup is repeated. The simulation results indicate that the F_d is robust against both types of non-normal errors. No difference in estimated size can be detected between the two non-normal cases and the normal case. The same result holds for F_{nd} when considering $t(6)$-distributed errors. On the other hand, errors from the $\chi^2(1) - 1$ distribution result in slightly higher rejection frequencies compared to the normal case at values of δ between -1 and about 0.6. The difference is the largest for the three smallest sample sizes, $T = 25, 50, 100$. For $T = 25$, the estimated size distortion is up to 3 percentage points higher, highest for values of δ close to -1. When $T = 100$ the size is only up to 1 percentage points higher. At other sample sizes, and for all sample sizes at values of δ close to 1, no effect on the size can be detected compared to the normal case.

Figure 1.3 Size of the F_d statistic for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.

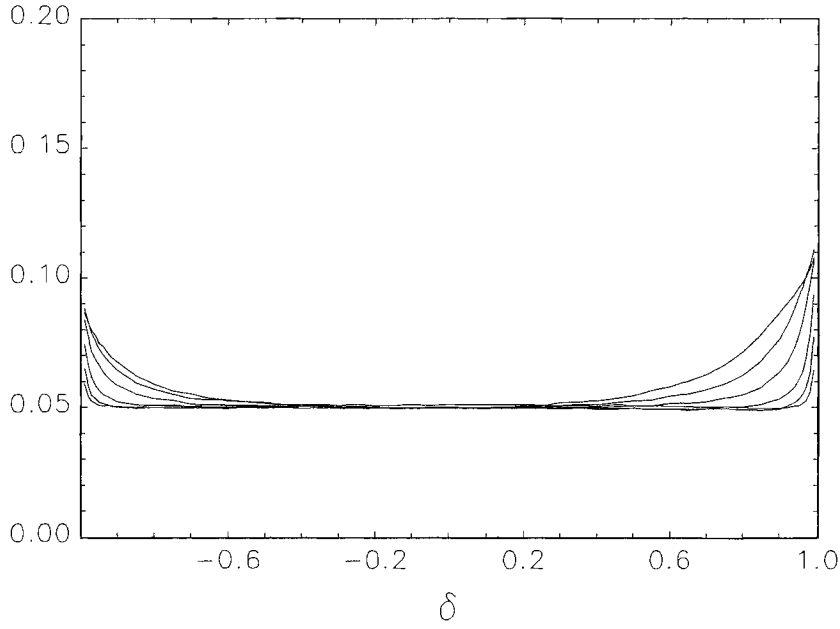
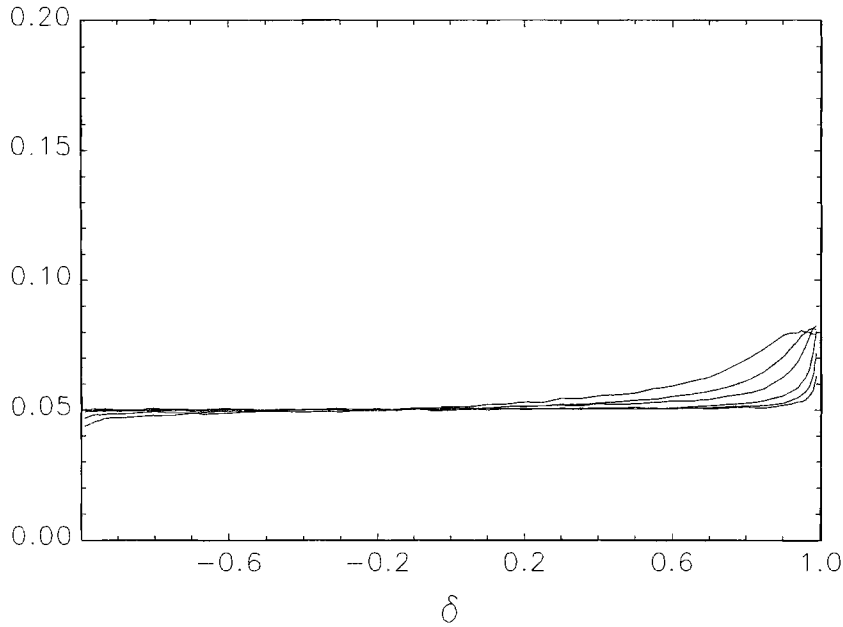


Figure 1.4 Size of the ADF_d statistic for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.



1.4.2 Bootstrapping the critical values

In this section, a simple bootstrap method is suggested in order to correct the size distortion affecting the tests. Basing inference on a bootstrap distribution can substantially improve the finite sample properties of various test statistics, since the bootstrap p-value converges to the true p-value of the test as the number of bootstrap replications increases. In practice, the bootstrap p-value is estimated by simulation. For a survey on bootstrapping time series see Li and Maddala (1996). Caner and Hansen (2001) suggested, when testing the unit root hypothesis against the TAR model, basing the inference on a bootstrap approximation of the limit distribution of the test statistic under the null.

As is seen from the previous results, the small sample distributions of the two F tests, F_{nd} and F_d, depend on the parameter δ. Since the model simplifies to $y_t = \delta\Delta y_{t-1} + y_{t-1} + \varepsilon_t$ under H_{01}, and to $y_t = \delta\Delta y_{t-1} + \alpha + y_{t-1} + \varepsilon_t$ under H_{02}, a model-based bootstrap can be used for estimating p-values for both tests. Consider first statistic F_{nd}, let $\widehat{\delta}$ and $\widehat{D}$ be estimates of δ and the distribution D of the errors ε_t. Let ε_t^b be a random draw from $\widehat{D}$, and generate

the bootstrap time series

$$y_t^b = \widehat{\delta}\Delta y_{t-1}^b + y_{t-1}^b + \varepsilon_t^b, \quad t = 1, \ldots, T. \tag{1.18}$$

Initial values for the resampling can be set to sample values of the demeaned series. The distribution of the series y_t^b is called the bootstrap distribution of the data. The test statistic, now called F_{nd}^b, is calculated from the resampled series y_t^b in the usual way. Repeating this resampling operation B times yields the empirical distribution of F_{nd}^b, which is the bootstrap distribution of F_{nd}, completely determined by $\widehat{\delta}$ and $\widehat{D}$. For a large number of independent F_{nd}^b tests, estimated from the B resampled series, the bootstrap p-value, defined by $p_T = P\left(F_{nd}^b > F_{nd}\right)$, can then be approximated by the frequency of simulated F_{nd}^b that exceeds the observed value of F_{nd}.

The resampling scheme is easily modified to fit statistic F_d. In order to obtain the bootstrap distribution of F_d, model (2.11) is augmented as follows:

$$y_t^b = \widehat{\delta}\Delta y_{t-1}^b + \widehat{\alpha} + y_{t-1}^b + \varepsilon_t^b, \quad t = 1, \ldots, T, \tag{1.19}$$

where $\widehat{\alpha}$ is the least squares estimator of α. The corresponding bootstrap distribution and the p-value are obtained as before.

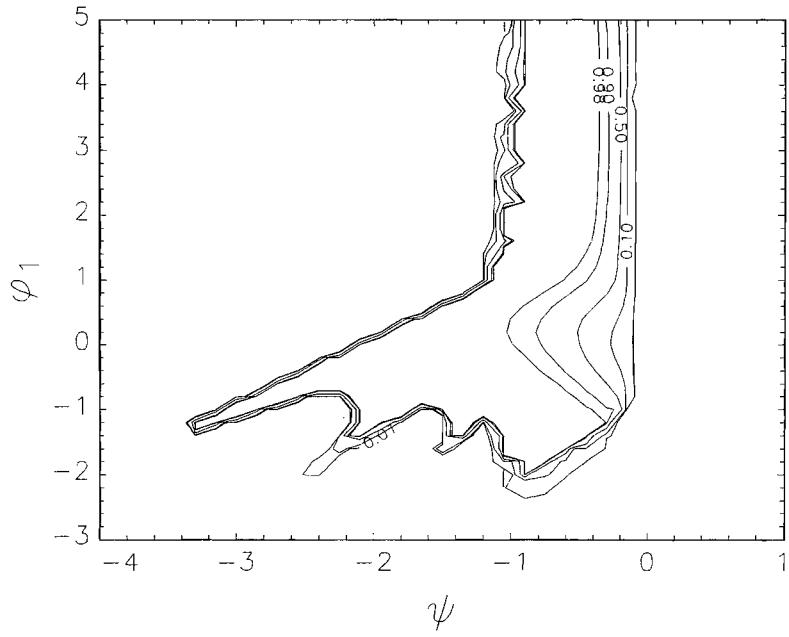
1.4.3 Power simulations

The power study involves generating data from a stationary STAR model under the alternative hypothesis. However, there is no analytical answer to the question of which parameter combinations correspond to a stationary model under the alternative hypothesis. Only a guideline can be accomplished by setting the transition function $F = 0$ or 1, in order to obtain reasonable boundaries to the model parameters. On the other hand, simulations can show where the model is nonstationary, at least in the sense that a realization of $\{y_t\}$ cross a predetermined boundary. Such a crossing is taken to mean that the model is nonstationary for that specific choice of parameters. Using this idea in the Monte Carlo setup, the time series y_t is said to be nonstationary if $|y_t| > \sigma t$ for $t > 1000000$, where σ equals the standard error of the errors ε_t in (1.1). This is of course just a rough indication on nonstationarity, especially for parameter choices very close to the boundary between the stationary and nonstationary regions.

Data is generated from the STAR model (1.1) using $T = 50$, assuming $\{\varepsilon_t\} \sim nid(0, 1)$. Since no extra power, compared to the ADF tests, can be expected if a regime shift only involves the intercept, the two constants θ_0 and ϕ_0 are set to zero. The size of the regime shift is then determined by the

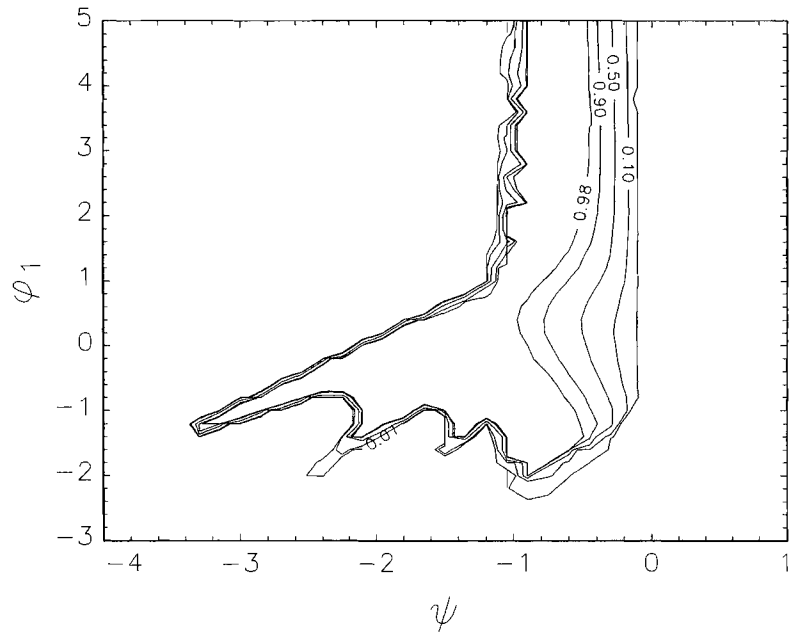
parameters θ_1 and φ_1. Let $\theta_1 = -\varphi_1$ for simplicity, and let the parameters in the logistic function be $\gamma = 10$ and $c = 0$. The model is then determined by only two unknown parameters, φ_1 and ψ. Figures 1.5, 1.6, 1.8 and 1.9 shows contour plots of the power to the four tests, F_{nd}, ADF_{nd}, F_d and ADF_d, for a grid of φ_1 and ψ values. The difference in power between F_{nd} and ADF_{nd}, expressed as power of F_{nd} minus power of ADF_{nd}, can be found in Figure 1.7. The difference in power between F_d and ADF_d, expressed in the similar fashion, is shown in Figure 1.10. Due to substantial computational costs, the number of replications only equals 10 000, and 500 bootstrap replications are used to estimate the p-values.

Figure 1.5 Power of the F_{nd} statistic for $T = 50$ observations.



The simulations show that the parameter combinations (φ_1, ψ) in the area outside the contour lines result in processes whose realizations grow without a bound with the number of observations. The largest gain in power from using F instead of ADFoccurs at parameter combinations of φ_1 and ψ such that $-1 < \psi < 0$. Other combinations indicate negligible or small differences in power between the two pairs of tests. The single largest gain for the F_{nd} test is as much as 56.7 percentage points whereas the smallest one equals -6.1 percentage points. The number of pairs (φ_1, ψ) with a negative gain corresponds to about 2.2% of the total number of pairs. The single largest and smallest

Figure 1.6 Power of the ADF_{nd} statistic for $T = 50$ observations.



gain for the F_d test are 52.2 and −6.9 percentage points, respectively. The gain is even here negative for about 2.2% of the pairs.

The negative gain for some of the parameter combinations can not be explained by sampling error alone. The main explanation is that the alternative STAR model at these parameter combinations is very close to the linear alternative model considered in the ADF test. The auxiliary model (1.4) is then very close to or indistinguishable from the ADF model, and the power of the tests is reduced because of the extra parameter to be tested. As a whole, the simulation results show, however, that the F_{nd} and F_d tests have about the same or considerable higher power than the corresponding standard ADF tests when the alternative exhibit nonlinear behavior.

Figure 1.7 Difference in power, $F_{nd} - ADF_{nd}$.

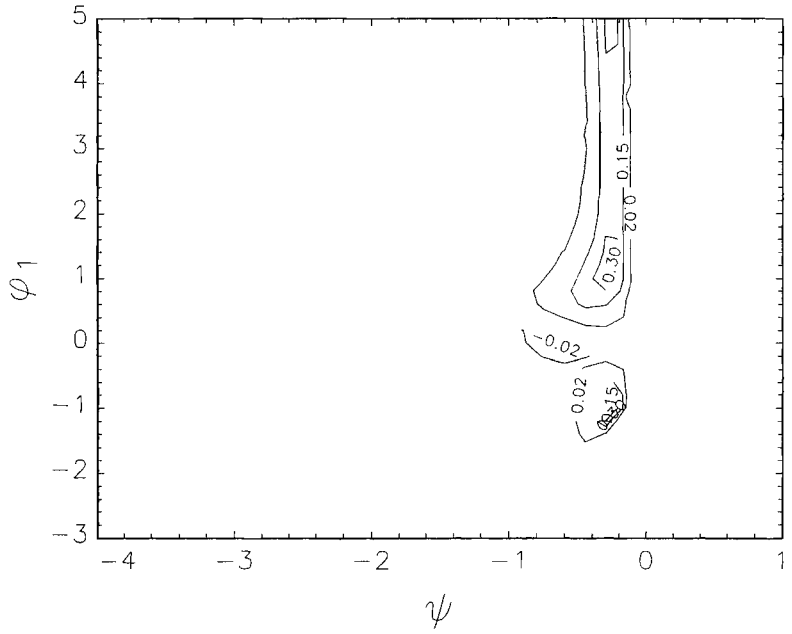


Figure 1.8 Power of the F_d statistic for $T = 50$ observations.

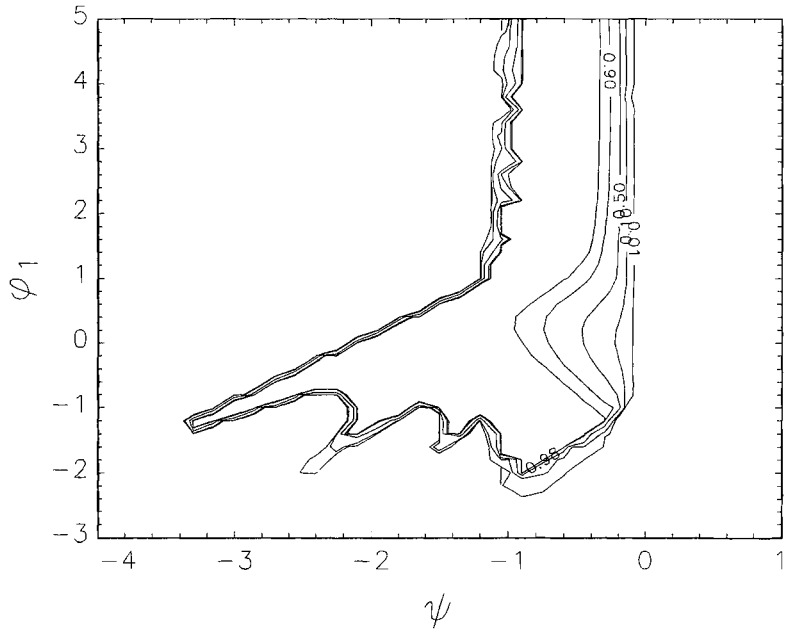


Figure 1.9 Power of the ADF_d statistic for $T = 50$ observations.

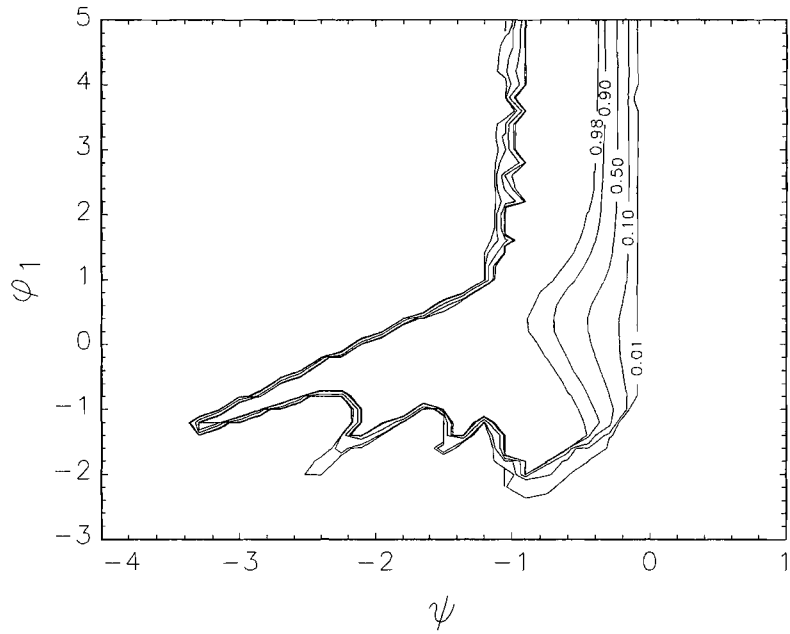
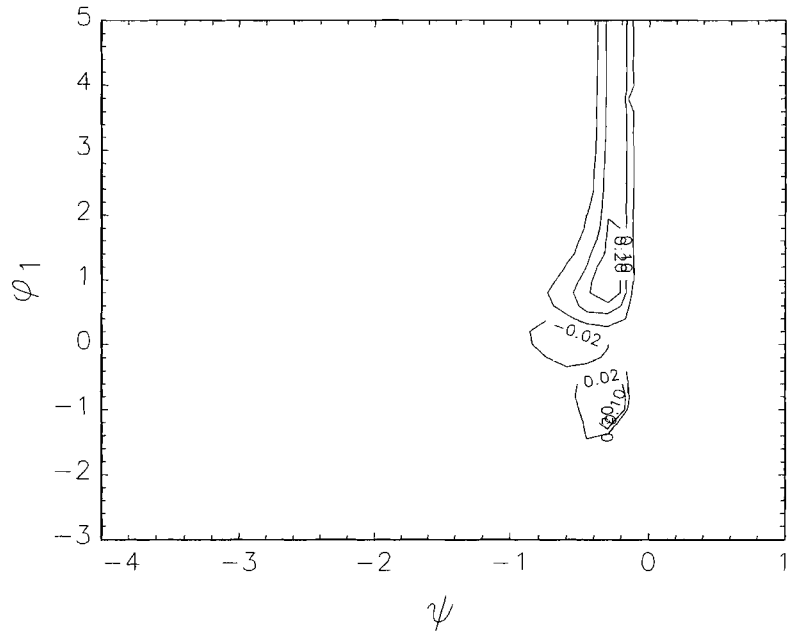


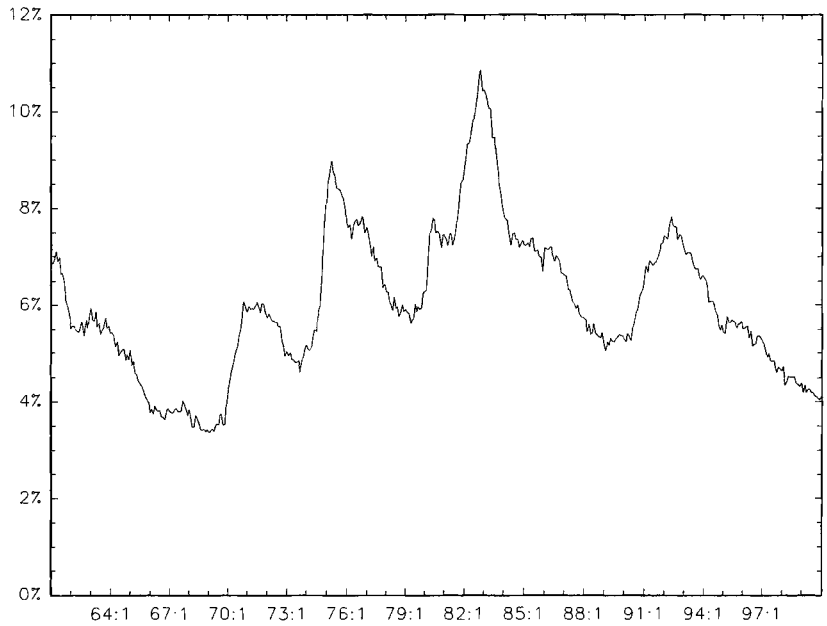
Figure 1.10 Difference in power, $F_d - ADF_d$.



1.5 Application

In order to demonstrate the use of the tests in practice they are applied to the seasonally adjusted monthly U.S. unemployment rate, from January 1961 to February 2000, obtained from SourceOECD. The sample period contains 468 observations of the differenced series. A plot of the time series can be found in Figure 1.11. A typical feature of the series is that there are periods of rapid increase of unemployment. An interesting feature is the asymmetry around the peaks, that is, the increase in unemployment is indeed more rapid than in the subsequent decrease. Such asymmetric behavior cannot be described properly with a linear model. Whether or not this unemployment rate can be assumed stationary is not quite clear from the figure, although a visual inspection may suggest mean reversion.

Figure 1.11 Seasonally adjusted unemployment for U.S.A. in %.



Carrying out the tests, using ordinary least squares, the auxiliary model (1.4) is estimated under the two null hypotheses H_{01} and H_{02}, and under the alternative hypothesis. The estimated equations and sums of squared residuals are as follows:

Under $H_{01} : \phi = \alpha = 0,\ \rho = 1,$

$$y_t = 0.1380\Delta y_{t-1} + y_{t-1} + \widehat{\varepsilon}_t, \qquad SSR_{01} = 13.49$$

Under $H_{02} : \phi = 0,\ \rho = 1,$

$$y_t = 0.1370\Delta y_{t-1} - 0.0051 + y_{t-1} + \widehat{\varepsilon}_t, \quad SSR_{02} = 13.48$$

Under the alternative hypothesis,

$$y_t = -0.5301\Delta y_{t-1} + 0.0999 y_{t-1}\Delta y_{t-1} + 0.0302 + 0.9938 y_{t-1} + \widehat{\varepsilon}_t, \qquad SSR = 13.09$$

Table 1.2 contains the observed values of the four test statistics F_{nd}, ADF_{nd}, F_d and ADF_d. The p-value for each test statistic has been estimated using the model-based bootstrap method with $B = 10\ 000$.

Table 1.2. Estimated test statistics and p-value.

	F_{nd}	ADF_{nd}	F_d	ADF_d
statistic	4.7554	0.9209	6.9160	1.4181
p-value	0.0174	0.8294	0.0095	0.6332

Since the p-values of F_{nd} and F_d are about 1.7% and 0.9%, the null hypotheses can be rejected at the 5% significance level. For 468 observations, the corresponding critical values from Table 1.1 are about 3.6 and 4.9 for the 5% level tests. The null hypotheses are thus also rejected when the inference is based on the critical values in Table 1.1. Using these critical values appears justified because of the negligible size distortion for values of δ close to the consistently estimated parameter $\widehat{\delta} = 0.1380$. The actual size for $\widehat{\delta}$ is very close to 5% as seen in Figures 1.1 and 1.3. These rejections support the conclusion that the U.S. seasonally adjusted monthly unemployment rate can be better characterized by a stationary nonlinear model than by a random walk. The use of the two ADF tests, however, leads to a different conclusion as the null hypothesis of a unit root is not rejected at any customary level of significance. The p-values are large, 83% and 63%, and corresponding critical values are 4.7 and 8.3 for ADF_{nd} and ADF_d respectively. As for the F tests, the actual size is very close to 5%, see Figures 1.2 and 1.4, and basing the ADF tests on the critical values in Table 1.1 would result in no or negligible size distortions.

1.6 Conclusions

This paper contains statistical theory for testing the unit root hypothesis against the smooth transition autoregressive (STAR) model. Some new limit

results, two F-type test statistics and critical values for them are presented. The joint hypothesis of unit root and linearity allows one to distinguish between random walk processes, with or without drift, and stationary nonlinear processes of the smooth transition autoregressive type. This is important in applications because steps taken in modelling the series are likely to be drastically different depending on whether or not the unit root hypothesis is rejected. For illustration, the tests are applied to the seasonally adjusted U.S. monthly unemployment rate. The unit root hypothesis is strongly rejected, giving support to the hypothesis that the unemployment series follows a STAR model rather than a random walk. The test result is of interest when the possible presence of hysteresis in the U.S. unemployment is considered.

Appendix A

Proof of Theorem 1

(a) Consider first

$$T^{-1}\sum_{t=1}^{T}\xi_t u_t^2 = T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^2 - Eu_t^2\right) + T^{-1}\sum_{t=1}^{T}\xi_t Eu_t^2. \tag{A.1}$$

Now, since $\xi_t = \xi_{t-1} - u_t$, the first sum on the right-hand side equals

$$T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^2 - Eu_t^2\right) = T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^2 - Eu_t^2\right) + T^{-1}\sum_{t=1}^{T}u_t\left(u_t^2 - Eu_t^2\right) \tag{A.2}$$

where the last term is $O_p(1)$, or $o_p(1)$ if $Eu_t^3 = 0$ as in the Gaussian case. Now let $\upsilon_t = \left(u_t, u_t^2 - Eu_t^2\right)'$, $V_t = \sum_{i=1}^{t}\upsilon_i$ and $V_0 = 0$. Then, from Hansen (1992), Theorem 4.1, it is known that the sum $T^{-1}\sum_{t=1}^{T}V_{t-1}\upsilon_t'$ converges weakly to a stochastic integral. Thus $T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^2 - Eu_t^2\right) = O_p(1)$ and

$$T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^2 = T^{-3/2}\sum_{t=1}^{T}\xi_t Eu_t^2 + o_p(1) \Rightarrow \gamma_0\lambda\int_0^1 W(r)\,dr. \tag{A.3}$$

(b) Using the same idea as in (a),

$$T^{-1}\sum_{t=1}^{T}\xi_t u_t^3 = T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^3 - Eu_t^3\right) + T^{-1}\sum_{t=1}^{T}\xi_t Eu_t^3. \tag{A.4}$$

The first sum on the right-hand side then equals

$$T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^3-Eu_t^3\right)=T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^3-Eu_t^3\right)+T^{-1}\sum_{t=1}^{T}u_t\left(u_t^3-Eu_t^3\right) \tag{A.5}$$

where the last term is $O_p(1)$. Let $v_t=\left(u_t,u_t^3-Eu_t^3\right)'$, $V_t=\sum_{i=1}^{t}v_i$ and $V_0=0$. Then, again from Hansen (1992), Theorem 4.1, the sum $T^{-1}\sum_{t=1}^{T}V_{t-1}v_t'$ converges weakly to a stochastic integral. We have $T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^3-Eu_t^3\right)=O_p(1)$ and

$$T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^3=T^{-3/2}Eu_t^3\sum_{t=1}^{T}\xi_t+o_p(1)\Rightarrow Eu_t^3\lambda\int_0^1 W(r)\,dr. \tag{A.6}$$

(*c*) As a starting-point, consider the sum

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^2=T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^2-Eu_t^2\right)+T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2Eu_t^2 \tag{A.7}$$

and let v_t, V_t and V_0 be as in (*a*). Let Assumption 1 hold with $\eta=3$. It then follows from Hansen (1992), Theorem 4.2, that the sum $T^{-3/2}\sum_{t=1}^{T}\left(V_{t-1}\otimes V_{t-1}\right)v_t'$ converges weakly to a stochastic integral. This implies that

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^2-Eu_t^2\right)=O_p(1) \tag{A.8}$$

so that

$$T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^2=T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2Eu_t^2+o_p(1)\Rightarrow\gamma_0\lambda^2\int_0^1 W^2(r)\,dr. \tag{A.9}$$

It remains to be shown that

$$T^{-2}\sum_{t=1}^{T}\xi_t^2u_t^2=T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^2\text{ as }T\longrightarrow\infty. \tag{A.10}$$

Since $\xi_t = \xi_{t-1} + u_t$,

$$\sum_{t=1}^{T} \xi_t^2 u_t^2 = \sum_{t=1}^{T} \left(\xi_{t-1} + u_t\right)^2 u_t^2 = \sum_{t=1}^{T} \xi_{t-1}^2 u_t^2 + 2\sum_{t=1}^{T} \xi_{t-1} u_t^3 + \sum_{t=1}^{T} u_t^4. \tag{A.11}$$

It follows directly from (A.11) that

$$\sum_{t=1}^{T} \left(\xi_t^2 - \xi_{t-1}^2\right) u_t^2 = 2\sum_{t=1}^{T} \xi_{t-1} u_t^3 + \sum_{t=1}^{T} u_t^4, \tag{A.12}$$

where the two sums on the right-hand side are $O_p\left(T^{3/2}\right)$ and $O_p\left(T\right)$ respectively. The difference is thus $o_p\left(T^2\right)$, implying that (A.9) holds.

(*d*) Let $v_t = (u_t, u_{t-1}\varepsilon_t)'$, $V_t = \sum_{i=1}^{t} v_i$ and $V_0 = 0$. Then, since $T^{-1/2}V_T \Longrightarrow \left(\lambda W\left(1\right), \sigma\sqrt{\gamma_0}W\left(1\right)\right)'$, from Hansen (1992), Theorem 4.1, it follows that the elements of the sum

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} V_{t-1}v_t' &= T^{-1}\sum_{t=1}^{T} \begin{bmatrix} \xi_{t-1} \\ \sum_{i=1}^{t-1} u_{i-1}\varepsilon_i \end{bmatrix} \begin{bmatrix} u_t & u_{t-1}\varepsilon_t \end{bmatrix} \\ &= T^{-1}\sum_{t=1}^{T} \begin{bmatrix} \xi_{t-1}u_t & \xi_{t-1}u_{t-1}\varepsilon_t \\ u_t\sum_{i=1}^{t-1} u_{i-1}\varepsilon_i & u_{t-1}\varepsilon_t\sum_{i=1}^{t-1} u_{i-1}\varepsilon_i \end{bmatrix} \end{aligned} \tag{A.13}$$

converge weakly to some stochastic integrals. In particular,

$$T^{-1}\sum_{t=1}^{T} \xi_{t-1}u_{t-1}\varepsilon_t \Rightarrow \sigma\sqrt{\gamma_0}\lambda \int_0^1 W\left(r\right) dB\left(r\right) + \Lambda_{1,2} \tag{A.14}$$

where $\Lambda_{1,2}$ is element $(1,2)$ in the matrix

$$\begin{aligned} \Lambda &= \lim_{T\longrightarrow\infty} \frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E \begin{bmatrix} u_i \\ u_{i-1}\varepsilon_i \end{bmatrix} \begin{bmatrix} u_j & u_{j-1}\varepsilon_j \end{bmatrix} \\ &= \lim_{T\longrightarrow\infty} \frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E \begin{bmatrix} u_iu_j & u_iu_{j-1}\varepsilon_j \\ u_ju_{i-1}\varepsilon_i & u_{i-1}u_{j-1}\varepsilon_j\varepsilon_i \end{bmatrix} \end{aligned} \tag{A.15}$$

Then, since $u_i u_{j-1}$ and ε_j are independent for all $j \geq i+1$,

$$\begin{aligned} \Lambda_{1,2} &= \lim_{T\longrightarrow\infty} \frac{1}{T} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E\left(u_i u_{j-1} \varepsilon_j\right) \\ &= \lim_{T\longrightarrow\infty} \frac{1}{T} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E\left(u_i u_{j-1}\right) E\left(\varepsilon_j\right) = 0. \end{aligned} \tag{A.16}$$

Thus the two Brownian motions $W(r)$ and $B(r)$ are independent, and the result follows.

(*e*) As a starting-point, consider the sum $T^{-3/2} \sum_{t=1}^{T} \xi_{t-1}^2 u_t$, and suppose that Assumption 1 holds for $\eta = 3$. Now let $v_t = u_t$, $V_t = \sum_{i=1}^{t} v_i$ and $V_0 = 0$. It follows from Hansen (1992), Theorem 4.2, that $T^{-3/2} \sum_{t=1}^{T} \left(V_{t-1} \otimes V_{t-1}\right) v_t$ converges weakly to a stochastic integral. Thus

$$T^{-3/2} \sum_{t=1}^{T} \xi_{t-1}^2 u_t \Rightarrow \lambda^3 \int_0^1 W^2(r)\, dW(r) + 2\lambda\Lambda \int_0^1 W(r)\, dr \tag{A.17}$$

where $\Lambda = \lim_{T\longrightarrow\infty} \frac{1}{T} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E u_i u_j$. The difference between the sum $\sum_{t=1}^{T} \xi_{t-1}^2 u_t$ and $\sum_{t=1}^{T} \xi_t^2 u_t$ is given by

$$\sum_{t=1}^{T} \left(\xi_t^2 - \xi_{t-1}^2\right) u_t = 2 \sum_{t=1}^{T} \xi_{t-1} u_t^2 + \sum_{t=1}^{T} u_t^3 \tag{A.18}$$

where the two sums on the right hand side are $O_p\left(T^{3/2}\right)$ and $O_p(T)$ respectively. This implies that

$$\sum_{t=1}^{T} \xi_t^2 u_t = O_p\left(T^{3/2}\right) = o_p\left(T^2\right), \tag{A.19}$$

which proves (*e*) and completes the proof of Theorem 1. ■

Appendix B

Proof of Theorem 2

Let $\Longrightarrow$ denote weak convergence.

(i) Since $u_t = \Delta y_t$ under the null hypothesis, the elements in the matrix equation (1.12) are

$$\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t x_t'\right)\Upsilon_T^{-1} = \tag{B.1}$$

$$= \begin{bmatrix} T^{-1}\sum_{t=1}^{T} u_{t-1}^2 & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1}^2 & T^{-1}\sum_{t=1}^{T} u_{t-1} & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1}^2 & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 u_{t-1}^2 & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 u_{t-1} \\ T^{-1}\sum_{t=1}^{T} u_{t-1} & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} & 1 & T^{-3/2}\sum_{t=1}^{T} y_{t-1} \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 u_{t-1} & T^{-3/2}\sum_{t=1}^{T} y_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}$$

and

$$\left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t\varepsilon_t\right)\right\}' = \tag{B.2}$$

$$= \begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} u_{t-1}\varepsilon_t & T^{-1}\sum_{t=1}^{T} y_{t-1}u_{t-1}\varepsilon_t & T^{-1/2}\sum_{t=1}^{T}\varepsilon_t & T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{bmatrix}.$$

Given the limit distributions in Theorem 1 and other known limit results, the ordinary least squares estimator, (1.12), converges weakly as follows

$$\Upsilon_T\left(b_T - \beta\right) = \left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t x_t'\right)\Upsilon_T^{-1}\right\}^{-1}\left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t\varepsilon_t\right)\right\} \tag{B.3}$$

$$\Rightarrow \begin{bmatrix} \gamma_0 & \gamma_0\lambda\int_0^1 W(r)\,dr & 0 & 0 \\ \gamma_0\lambda\int_0^1 W(r)\,dr & \gamma_0\lambda^2\int_0^1 W^2(r)\,dr & 0 & 0 \\ 0 & 0 & 1 & \lambda\int_0^1 W(r)\,dr \\ 0 & 0 & \lambda\int_0^1 W(r)\,dr & \lambda^2\int_0^1 W^2(r)\,dr \end{bmatrix}^{-1}$$
$$\times \begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\} \end{bmatrix}$$

$$= \frac{1}{\Sigma}\begin{bmatrix} \gamma_0^{-1}\int_0^1 W^2(r)\,dr & -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & 0 & 0 \\ -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & (\gamma_0\lambda^2)^{-1} & 0 & 0 \\ 0 & 0 & \int_0^1 W^2(r)\,dr & -\lambda^{-1}\int_0^1 W(r)\,dr \\ 0 & 0 & -\lambda^{-1}\int_0^1 W(r)\,dr & \lambda^{-2} \end{bmatrix}$$
$$\times \begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\} \end{bmatrix}$$

where $\Sigma = \int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2$. It then follows that, for the hypothesis $H_{01}: R\beta = r$ where $R = \begin{bmatrix} \mathbf{0} & I_3 \end{bmatrix}$, the F_{nd} test statistic defined by (1.13), equals

$$F_{nd} = (b_T-\beta)'(R\Upsilon_T)'\left\{s_T^2 R\Upsilon_T\left(\sum_{t=1}^T x_t x_t'\right)^{-1}\Upsilon_T R'\right\}^{-1} R\Upsilon_T(b_T-\beta)/3 \quad \text{(B.4)}$$

$$\Rightarrow \frac{1}{3\sigma^2\Sigma}\begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) & \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) & \sigma W(1) & \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\} \end{bmatrix}$$

$$\times \begin{bmatrix} \gamma_0^{-1}\left(\int_0^1 W(r)\,dr\right)^2 & -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & 0 & 0 \\ -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & (\gamma_0\lambda^2)^{-1} & 0 & 0 \\ 0 & 0 & \int_0^1 W^2(r)\,dr & -\lambda^{-1}\int_0^1 W(r)\,dr \\ 0 & 0 & -\lambda^{-1}\int_0^1 W(r)\,dr & \lambda^{-2} \end{bmatrix}$$

$$\times \begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\} \end{bmatrix}$$

$$\begin{aligned} = \frac{1}{3\Sigma}\Bigg[& W^2(1)\left(\int_0^1 W(r)\,dr\right)^2 - 2W(1)\int_0^1 W(r)\,dr\int_0^1 W(r)\,dB(r) \\ & + \left(\int_0^1 W(r)\,dB(r)\right)^2 + W^2(1)\int_0^1 W^2(r)\,dr \\ & - W(1)\int_0^1 W(r)\,dr\left\{W^2(1)-1\right\} + \frac{1}{4}\left\{W^2(1)-1\right\}^2\Bigg] \\ = \frac{1}{3}W^2(1) & + \frac{1}{3\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \int_0^1 W(r)\,dB(r)\right)^2 \\ & + \frac{1}{3\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \frac{1}{2}\left\{W^2(1)-1\right\}\right)^2 \end{aligned}$$

which ends the proof of (i).

(ii) Hypothesis H_{02} has the alternative representation $H_{02}: R\beta = r$ where $R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $r = \begin{bmatrix} 0 & 1 \end{bmatrix}'$. The F test statistic, defined by (1.13), equals

$$F_d = (b_T - \beta)'(R\Upsilon_T)'\left\{s_T^2 R\Upsilon_T\left(\sum_{t=1}^T x_t x_t'\right)^{-1}\Upsilon_T R'\right\}^{-1} R\Upsilon_T(b_T-\beta)/2 \tag{B.5}$$

and converges weakly as follows:

$$F_d \Rightarrow \frac{1}{2\sigma^2\Sigma}\left[\begin{array}{cccc} \sigma\sqrt{\gamma_0}W(1) & \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) & \sigma W(1) & \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\}\end{array}\right]$$

$$\times\left[\begin{array}{cccc} \gamma_0^{-1}\left(\int_0^1 W(r)\,dr\right)^2 & -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & 0 & 0 \\ -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & (\gamma_0\lambda^2)^{-1} & 0 & 0 \\ 0 & 0 & \left(\int_0^1 W(r)\,dr\right)^2 & -\lambda^{-1}\int_0^1 W(r)\,dr \\ 0 & 0 & -\lambda^{-1}\int_0^1 W(r)\,dr & \lambda^{-2}\end{array}\right]$$

$$\times\left[\begin{array}{c} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\}\end{array}\right]$$

$$\begin{aligned} = & \frac{1}{2\Sigma}\left(2W^2(1)\left(\int_0^1 W(r)\,dr\right)^2 - 2W(1)\int_0^1 W(r)\,dr\int_0^1 W(r)\,dB(r)\right. \\ & \left. + \left(\int_0^1 W(r)\,dB(r)\right)^2 - W(1)\int_0^1 W(r)\,dr\left\{W^2(1)-1\right\} + \frac{1}{4}\left\{W^2(1)-1\right\}^2\right) \\ = & \frac{1}{2\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \int_0^1 W(r)\,dB(r)\right)^2 \\ & + \frac{1}{2\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \frac{1}{2}\left\{W^2(1)-1\right\}\right)^2 \end{aligned}$$

where $\Sigma = \int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2$ as in (i). This completes the proof of (ii) and Theorem 2. ■

Bibliography

[1] Berben, R.-P., and van Dijk, D., (1999), "Unit root tests and asymmetric adjustment: A reassessment", Working paper, Erasmus University Rotterdam.

[2] Caner, M., and Hansen, B. E., (2001), "Threshold autoregression with a unit root", *Econometrica*, 69, 1555-1596.

[3] Diebold, F. X., and Rudebusch, G. D., (1990), "On the power of Dickey-Fuller tests against fractional alternatives", *Economics Letters*, 35, 155-160.

[4] Enders, W., and Granger, C. W. J., (1998), "Unit-root tests and asymmetric adjustment with an example using the term structure of interest rate", *Journal of Business and Economic Statistics*, 16, 304-311.

[5] Granger, C. W. J., and Teräsvirta, T., (1993), *Modelling nonlinear economic relationships*, Oxford, Oxford University Press.

[6] Hamori, S., and Tokihisa, A., (1997), "Testing for a unit root in the presence of a variance shift", *Economics Letters*, 57, 245-253.

[7] Hansen, B. E., (1992), "Convergence to stochastic integrals for dependent heterogeneous processes", *Econometric Theory*, 8, 489-500.

[8] Kapetanios, G., and Shin, Y. (2000), "Testing for a linear unit root against nonlinear threshold stationarity", Discussion paper, http://www.econ.ed.ac.uk/Academic_00Discuss.htm, University of Edingburgh.

[9] Kapetanios, G., Shin, Y., and Snell, A., (2003), "Testing for a unit root in the nonlinear STAR framework", *Journal of Econometrics*, 112, 359-379.

[10] Leybourne, S. J., Mills, T. C., and Newbold, P., (1998), "Spurious rejections by Dickey-Fuller tests in the presence of a break under the null", *Journal of Econometrics*, 87, 191-203.

[11] Li, H. and Maddala, G. S., (1996), "Bootstrapping time series models", *Econometric Reviews*, 15, 115-158.

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

[13] Nelson, C. R., Piger, J., and Zivot, E., (2001), "Markov regime switching and unit root tests", *Journal of Business and Economic Statistics*, 19, 404-415.

[14] Nelson, C. R., and Plosser, C. I., (1982), "Trends and random walks in macroeconomic time series", *Journal of Monetary Economics*, 10, 139-162.

[15] Perron, P., (1989), "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 57, 1361-1401.

[16] Pippenger, M. K., and Goering, G. E., (1993), "A note on the empirical power of unit root tests under threshold processes", *Oxford Bulletin of Economics and Statistics*, 55, 473-481.

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

[18] Teräsvirta, T., (1994b), "Testing linearity and modelling nonlinear time series", *Kybernetika*, 30, 319-330.

[19] Teräsvirta, T., (1998), "Modeling economic relationships with smooth transition regressions", in A. Ullah and D. E. A. Giles (eds.), *Handbook of applied economic statistics*, Dekker, New York, 507-552.

[20] Tong, H., (1983), *Threshold models in non-linear time series analysis*, New York: Springer-Verlag.

Chapter 2

A nonlinear alternative to the unit root hypothesis

2.1 Introduction

There are situations in which the standard Dickey-Fuller tests do not function as well as one might desire. Difficulties in detecting structural change, shifts in mean or growth rate, or nonlinear behavior have been noticed in several studies. Nelson and Plosser (1982) showed for a number of financial and macroeconomical time series that the Dickey-Fuller tests are unable to reject the unit root hypothesis. Pippenger and Goering (1993) argued that examining long-run economic relationships using the unit root tests is questionable in the presence of transaction costs or hysteresis thresholds. Perron (1989) argued that low power against structural breaks in level and growth rate can result in overstating the evidence in favor of unit roots. However, the standard Dickey-Fuller tests are also affected by size distortions in a number of situations. Size distortions, in the form of too frequent rejections of the null, have been observed when there is a single structural break in trend or variance under the null hypothesis, see Leybourne, Mills and Newbold (1998) and Hamori and Tokihisa (1997). Similar size distortion, when the true null model contains Markov regime switching in trend growth rate, was demonstrated by Nelson, Piger and Zivot (2001). They also showed power loss when the unit root hypothesis was tested against a true alternative process with a Markov-switching trend.

The emphasis in these studies, and in a major part of the early literature, has been on the linear model which today is increasingly viewed to be somewhat inadequate. Any possible nonlinear characteristics of the time series have thus been ignored. The increasing empirical evidence on nonlinear relationships and features in economical time series during the last few years has resulted in efforts to incorporate nonlinear models and techniques into the existing econometric framework. The literature on testing the unit root hypothesis against nonlinear models, or vice versa, has recently grown rapidly in this direction.

One of the more recent studies in this area is Caner and Hansen (2001) who analyzed and provided tests of the unit root hypothesis against the threshold autoregressive (TAR) model. The authors also proposed a bootstrap method to approximate the limit distribution of the test under the null, and showed that the unit root hypothesis can be rejected for the U.S. unemployment rate in favor of the nonlinear TAR model. Also considering the TAR model as an alternative to the unit root, Enders and Granger (1998) found that movements toward long-run equilibrium relationship of an interest rate are best described as an asymmetric process. Modifying the test in Enders and Granger (1998), Berben and van Dijk (1999) found evidence of asymmetric adjustments to-

wards long-run equilibrium for a number of forward premium time series. Kapetanios and Shin (2000) developed and analyzed the unit root test with the self-exciting threshold autoregressive (SETAR) model as the alternative process. Designed to take into account the threshold nature under the alternative, they reported some gain in power compared to the Dickey-Fuller test. Other studies have integrated the unit root test with nonlinear models with a smooth transition between the regimes. Kapetanios, Shin and Snell (2003) proposed a test of the joint unit root and linearity hypothesis against a very simple exponential smooth transition autoregressive (ESTAR) model that only allows a regime shift in the slope parameter. They were able to reject the unit root for a number of the real interest rates in favor of the ESTAR model. Bec, Salem and Carrasco (2002) tested the unit root hypothesis against a nonlinear logistic STAR model with three regimes. Rejecting the unit root for eleven out of twenty eight real exchange rates considered, their empirical results lent support to the so called purchasing power parity (PPP) hypothesis and indicated also strong mean reversion for large departures from PPP. Eklund (2003) proposed and analyzed the unit root test against the logistic smooth transition autoregressive (LSTAR) model, allowing for regime shift in both intercept and growth rate, and showed that the U.S. monthly unemployment rate is better described by a STAR model rather than a random walk.

In this paper, the recent work in Eklund (2003) is extended. Two tests are constructed for the joint linearity and unit root hypothesis against the second-order logistic STAR model. Compared to the earlier work in Eklund (2003), the alternative STAR model in this paper allows for regime shifts in intercept, growth rate and in level. The nonlinear model considered allow the adjustment towards long-run equilibrium to be sudden as well as smooth.

The paper is outlined as follows. In Section 2 the model is specified. Limit results and critical values are found in Section 3, while Section 4 includes a Monte Carlo study of the size and the power properties. Section 5 includes a small introduction to the so called purchasing power parity problem, and an empirical application on real exchange rates. Concluding remarks are given in Section 6, and mathematical proofs are presented in the appendix.

2.2 Model and joint unit root and linearity hypothesis

Consider the univariate smooth transition autoregressive (STAR) model

$$\Delta y_t = \theta_0+\theta_1\Delta y_{t-1}+\psi_1 y_{t-1}+(\varphi_0 + \varphi_1\Delta y_{t-1} + \psi_2 y_{t-1})\, F(\gamma, c_1, c_2, \Delta y_{t-1})+\varepsilon_t, \tag{2.1}$$

where Δy_t, and errors, ε_t, are assumed to be stationary, satisfying $E\varepsilon_t = 0$, $E|\varepsilon_t|^{6+r} < \infty$ for some $r > 0$, and $t = 1, \ldots, T$. The nonlinearity is introduced via the transition function $F(\cdot)$ which is a bounded continuous function such that $F(\cdot) \in [-\frac{1}{2}, \frac{1}{2}]$. This allows the dynamic behavior of Δy_t to change smoothly and nonlinearly with the transition variable Δy_{t-1} between the two regimes, $F(\cdot) = -\frac{1}{2}$ and $F(\cdot) = 1/2$. Several possibilities exists for the choice of the function $F(\cdot)$, see Granger and Teräsvirta (1993) and Teräsvirta (1998) for a detailed presentation and discussion. This paper will focus on (2.1) with a second-order logistic function

$$F(\gamma, c_1, c_2, \Delta y_{t-1}) = (1 + \exp[-\gamma(\Delta y_{t-1} - c_1)(\Delta y_{t-1} - c_2)])^{-1} - \frac{1}{2}, \tag{2.2}$$

where the parameters c_1 and c_2 are the threshold parameters and γ is the speed of transition between the regimes, $\gamma > 0$ for identification reasons. Note that the function $F(\cdot)$ is constant for $\gamma = 0$, in which case model (2.1) is linear. This fact can be used when testing linearity. However, testing the hypothesis $H_0 : \gamma = 0$ in model (2.1) cannot be performed directly, since this restriction involves an identification problem, see Luukkonen, Saikkonen and Teräsvirta (1988), Teräsvirta (1994a,b), Lin and Teräsvirta (1994) for details.

Applying the idea by Luukkonen et al. (1988), the identification problem can be circumvented by a first-order Taylor approximation of $F(\gamma, c_1, c_2, \Delta y_{t-1})$ around $\gamma = 0$. Inserting the approximation in (2.1) results in the following auxiliary model:

$$\begin{aligned}
\Delta y_t &= \theta_0 + \theta_1\Delta y_{t-1} + \psi_1 y_{t-1} \\
&\quad + (\varphi_0 + \varphi_1\Delta y_{t-1} + \psi_2 y_{t-1})\frac{\gamma}{4}(\Delta y_{t-1} - c_1)(\Delta y_{t-1} - c_2) + \varepsilon_t^* \\
&= \left(\theta_1 - \tfrac{\varphi_0\gamma(c_1+c_2)}{4} + \tfrac{\varphi_1\gamma c_1 c_2}{4}\right)\Delta y_{t-1} + \tfrac{\varphi_0\gamma-\varphi_1\gamma(c_1+c_2)}{4}(\Delta y_{t-1})^2 \\
&\quad + \frac{\varphi_1\gamma}{4}(\Delta y_{t-1})^3 - \frac{\psi_2\gamma(c_1+c_2)}{4} y_{t-1}\,\Delta\, y_{t-1} + \frac{\psi_2\gamma}{4} y_{t-1}(\Delta y_{t-1})^2 \\
&\quad + \left(\theta_0 + \frac{\varphi_0\gamma c_1 c_2}{4}\right) + \left(\psi_1 + \frac{\psi_2\gamma c_1 c_2}{4}\right) y_{t-1} + \varepsilon_t^*
\end{aligned}$$

$$\begin{aligned} = \; & \delta_1 \Delta y_{t-1} + \delta_2 (\Delta y_{t-1})^2 + \delta_3 (\Delta y_{t-1})^3 + \phi_1 y_{t-1} \Delta y_{t-1} \\ & + \phi_2 y_{t-1} (\Delta y_{t-1})^2 + \alpha + \zeta y_{t-1} + \varepsilon_t^*. \end{aligned} \tag{2.3}$$

where $\varepsilon_t^* = \varepsilon_t + (\varphi_0 + \varphi_1 \Delta y_{t-1} + \psi_2 y_{t-1}) R_1 (\gamma, \Delta y_{t-1})$, and R_1 is the remainder. Moving y_{t-1} from the left-hand to the right-hand side in (2.3) yields the form

$$\begin{aligned} y_t = \; & \delta_1 \Delta y_{t-1} + \delta_2 (\Delta y_{t-1})^2 + \delta_3 (\Delta y_{t-1})^3 + \phi_1 y_{t-1} \Delta y_{t-1} \\ & + \phi_2 y_{t-1} (\Delta y_{t-1})^2 + \alpha + \rho y_{t-1} + \varepsilon_t^*, \end{aligned} \tag{2.4}$$

where the original linearity condition, $\gamma = 0$, now corresponds to $\delta_2 = \delta_3 = \phi_1 = \phi_2 = 0$. Note that $\varepsilon_t^* = \varepsilon_t$ under the linearity hypothesis, since $R_1 = 0$ when $\gamma = 0$. As the regression model for the Augmented Dickey-Fuller (ADF) test, with a constant and one lag of Δy_t, is nested in this auxiliary model, a joint test of the linearity and the unit root hypothesis amounts to testing the hypothesis $H_{01} : \delta_2 = \delta_3 = \phi_1 = \phi_2 = \alpha = 0$, $\rho = 1$ in (2.4). Under this hypothesis equation (2.4) becomes

$$y_t = \delta_1 \Delta y_{t-1} + y_{t-1} + \varepsilon_t. \tag{2.5}$$

In reduced form equation (2.5) equals

$$\Delta y_t = \frac{\varepsilon_t}{1 - \delta_1 L} = \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} = \omega (L) \varepsilon_t = u_t, \tag{2.6}$$

where L is the lag operator, i.e. $L y_t = y_{t-1}$. Under H_{01}, $\{y_t\}$ is thus a unit root process without drift. Note that if H_{01} is rejected and the alternative is accepted as a basis for further modelling, the parameters θ_0 and ψ_1 should be included in the alternative model. This follows from the fact that testing $\alpha = 0$ also implies a test of $\theta_0 = 0$ in equation (2.3) under the original linearity condition $\gamma = 0$. The same reason holds for ψ_1. Excluding $\alpha = 0$ from H_{01} results in another null hypothesis H_{02} that allows for a unit root with a drift component. Assuming that Δy_t is stationary implies that δ_1 is restricted to $|\delta_1| < 1$ in equation (2.5). For $\delta_1 = 1$, Δy_t is $I(1)$ so that y_t is $I(2)$. Furthermore, when $\delta_1 = -1$, y_t has a negative unit root, and values $|\delta_1| > 1$ implies a nonstationary Δy_t process. Thus, as a consequence, problems can arise in practice if the value of δ_1 is near -1 or 1, as such values may distort the size of the tests.

2.3 Limit results and critical values

In this section, the necessary results for the asymptotic theory for testing H_{01} and H_{02} is derived. The least squares estimator $b_T = \left(\widehat{\delta}_1, \widehat{\delta}_2, \widehat{\delta}_3, \widehat{\phi}_1, \widehat{\phi}_2, \widehat{\alpha}, \widehat{\rho}\right)'$ of the parameters in the auxiliary model (2.4) has the form

$$b_T - \beta = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t, \tag{2.7}$$

where $x_t = \left(\Delta y_{t-1}, (\Delta y_{t-1})^2, (\Delta y_{t-1})^3, y_{t-1}\Delta y_{t-1}, y_{t-1}(\Delta y_{t-1})^2, 1, y_{t-1}\right)'$. In order to derive the asymptotic limit distributions for the elements of (2.7) not previously considered in the literature, the following assumption, see Hansen (1992), is assumed to be satisfied.

Assumption 1 *For some $q > \eta > 2$, $\{v_t\}$ is a zero mean, strong mixing sequence with mixing coefficients of size $-q\eta/(q-\eta)$, and $\sup_{t\geq 1} \|v_t\| = C < \infty$. In addition,*

$$T^{-1}E\left(V_T V_T'\right) \longrightarrow \Omega < \infty \text{ as } T \longrightarrow \infty, \text{ where } V_T = \sum_{t=1}^{T} v_t.$$

Allowing for weakly dependent heterogeneous data, this assumption and the theorems by Hansen (1992) are applicable to a number of different processes that typically arise in econometric applications. Assumption 1 will be used in the following theorem that contains a set of new convergence results needed in this work.

Theorem 1 *Let u_t, defined in (2.6), satisfy Assumption 1, and let $\{\varepsilon_t\}$ be an i.i.d. sequence with mean zero, variance σ^2, and $E|\varepsilon_t|^{6+r} < \infty$ for some $r > 0$. Define*

$$\begin{aligned}
&\gamma_j = E(u_t u_{t-j}) = \sigma^2 \sum_{s=0}^{\infty} \omega_s \omega_{s+j} && , j = 0, 1, \ldots \\
&\mu_j = E u_t^j && , j = 3, 4, \ldots \\
&\lambda = \sigma \sum_{j=0}^{\infty} \omega_j = \sigma\omega(1) && \\
&\xi_t = \sum_{i=0}^{t} u_i && , t = 1, 2, \ldots, T,
\end{aligned}$$

with $\xi_0 = 0$. Then the following sums converge jointly

$$(a)\quad T^{-1/2}\sum_{t=1}^{T} u_{t-1}^{2}\varepsilon_t \;\Rightarrow\; \sigma\sqrt{\mu_4}W(1)$$

$$(b)\quad T^{-1/2}\sum_{t=1}^{T} u_{t-1}^{3}\varepsilon_t \;\Rightarrow\; \sigma\sqrt{\mu_6}W(1)$$

$$(c)\quad T^{-1}\sum_{t=1}^{T} \xi_{t-1}u_{t-1}^{2}\varepsilon_t \;\Rightarrow\; \sigma\sqrt{\mu_4}\lambda\int_0^1 W(r)\,dB(r)$$

$$(d)\quad T^{-3/2}\sum_{t=1}^{T} \xi_t u_t^{4} \;\Rightarrow\; \mu_4\lambda\int_0^1 W(r)\,dr$$

$$(e)\quad T^{-3/2}\sum_{t=1}^{T} \xi_t u_t^{5} \;\Rightarrow\; \mu_5\lambda\int_0^1 W(r)\,dr$$

$$(f)\quad T^{-2}\sum_{t=1}^{T} \xi_t^{2} u_t^{3} \;\Rightarrow\; \mu_3\lambda^2\int_0^1 W^2(r)\,dr$$

$$(g)\quad T^{-2}\sum_{t=1}^{T} \xi_t^{2} u_t^{4} \;\Rightarrow\; \mu_4\lambda^2\int_0^1 W^2(r)\,dr$$

where $W(r)$ and $B(r)$ are two independent standard Brownian motions defined for $r \in [0,1]$.

Proof See the appendix.

Consulting the rates of convergence in Theorem 1, some recent limit results from Eklund (2003), and other known results one can define the following scaling matrix

$$\Upsilon_T = diag\left(\; T^{1/2},\;\; T^{1/2},\;\; T^{1/2},\;\; T,\;\; T,\;\; T^{1/2},\;\; T\;\right). \tag{2.8}$$

Pre-multiplying both sides of (2.7) by the scaling matrix Υ_T, finite limits to the rescaled ordinary least squares estimates are given by

$$\Upsilon_T\left(b_T-\beta\right) = \left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t x_t'\right)\Upsilon_T^{-1}\right\}^{-1}\left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t\varepsilon_t\right)\right\}. \tag{2.9}$$

Now write $H_{01}: R\beta = r$, where $R = \left[\begin{array}{cc}\mathbf{0} & I_6\end{array}\right]$, $\mathbf{0}$ is a (6×1) column vector, $\beta = (\delta_1,\delta_2,\delta_3,\phi_1,\phi_2,\alpha,\rho)'$, and $r = \left(\begin{array}{cccccc}0 & 0 & 0 & 0 & 0 & 1\end{array}\right)'$. An F test statistic can then be defined in the usual way as

$$F = \left(b_T-\beta\right)'\left(R\Upsilon_T\right)'\left\{s_T^2 R\Upsilon_T\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}\Upsilon_T R'\right\}^{-1} R\Upsilon_T\left(b_T-\beta\right)/k, \tag{2.10}$$

where $k = 6$ equals the number of restrictions, and s_T^2 is a consistent estimator of the residual variance in (2.4). The test statistic for hypothesis H_{02} that allows for the presence of a drift component can be defined similarly setting $k = 5$ and excluding row 5 of R and r. The two resulting test statistics will be called F_{nd} and F_d, where nd and d corresponds to '*no drift*' and '*drift*' respectively. Clearly, from Theorem 1, these test statistics do not have standard asymptotic distributions, as would be the case for stationary processes.

As the analytical limit expressions to the F test in (2.10) only can be computed with considerably difficulty, explicit expressions for F_{nd} and F_d are not given. The main reason for this difficulty is the problem of first obtaining and then simplifying the expressions of the inverses of the relatively large matrices that appear in equation (2.10): their dimensions are (5×5), (6×6) and (7×7).

Critical values for F_{nd} and F_n can, however, easily be obtained by a Monte Carlo simulation. Generating data from the null model (2.5) for $\delta_1 = 0$, Table 2.1 contains the critical values based on 1000000 replications for these statistics corresponding to sample sizes $T = 25, 50, 100, 250, 500, 5000$. Since the explicit limit expressions of the statistics are not known, critical values cannot be calculated for the asymptotic null distributions.

Table 2.1 Critical values for the test statistics F_{nd} and F_d, when $\delta_1 = 0$.

	F_{nd}					F_d				
T	0.10	0.05	0.025	0.01	0.001	0.10	0.05	0.025	0.01	0.001
25	2.49	3.06	3.65	4.49	7.00	2.70	3.36	4.05	5.02	7.85
50	2.28	2.71	3.12	3.68	5.18	2.50	3.00	3.50	4.15	5.84
100	2.22	2.60	2.97	3.43	4.59	2.45	2.90	3.33	3.90	5.28
250	2.20	2.56	2.90	3.34	4.37	2.44	2.86	3.27	3.78	5.05
500	2.20	2.55	2.89	3.31	4.33	2.44	2.86	3.27	3.77	4.99
5000	2.20	2.55	2.88	3.30	4.29	2.44	2.86	3.27	3.77	4.96

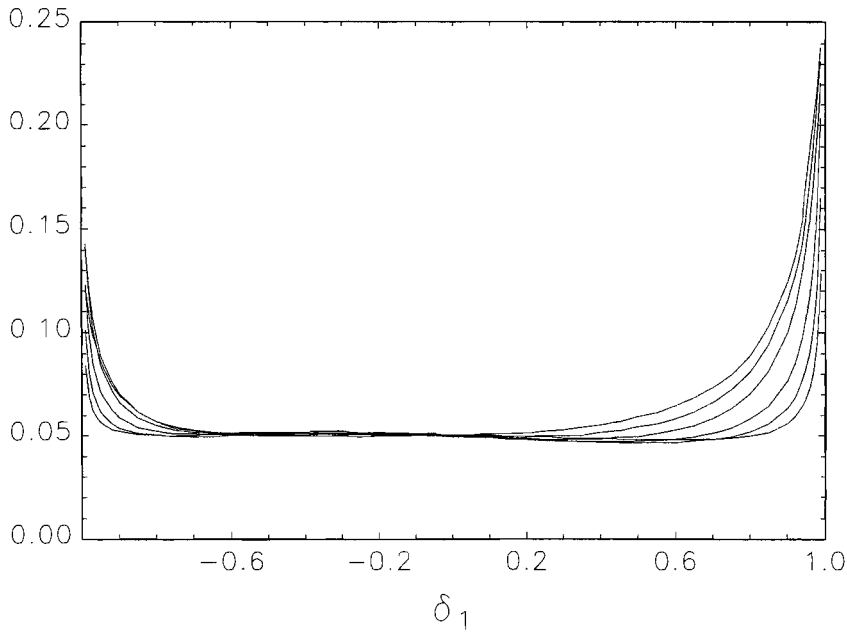
2.4 Small sample properties of the tests

In this section the size and the power of F_{nd}, F_d and the two corresponding ADF tests are compared. The latter are called ADF_{nd} and ADF_d respectively. A simple method to estimate p-values of the tests is also proposed in order to adjust for size distortion that is present in the tests for values of δ close to -1 or 1.

2.4.1 Size simulations

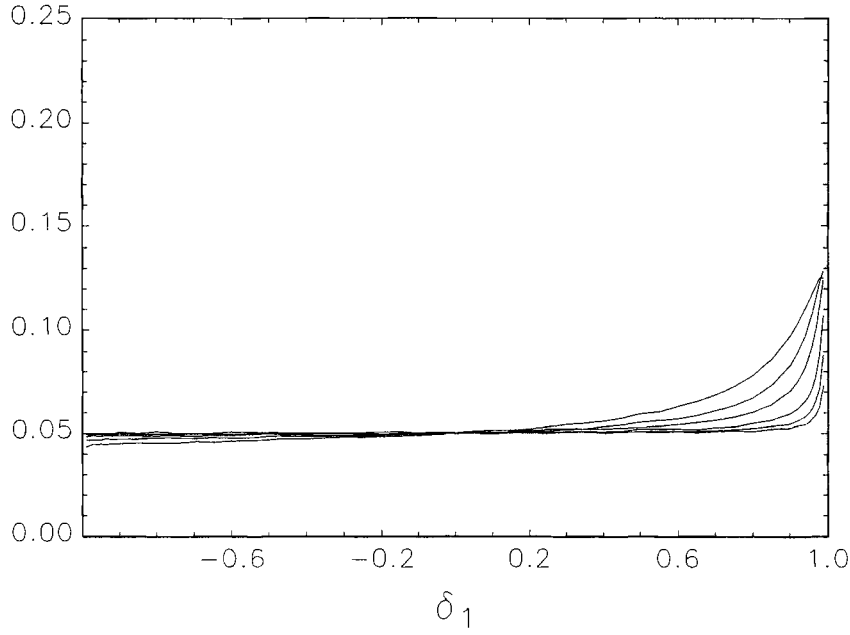
In order to consider the size of the tests, data are generated from model (2.5) under the null hypothesis, assuming $\{\varepsilon_t\} \sim nid\,(0,1)$. Using Monte Carlo simulations with 1000000 replications and critical values from Table 2.1, the rejection frequencies are calculated for the sample sizes $T = 25, 50, 100, 250, 500, 1000$. The nominal size for each test equals 5%. Since the null model depends on the parameter δ_1 in small samples, the size have been calculated for a number of different values of δ_1 ranging from -1 to 1. In the ADF tests, the correct number of lags (one) is assumed known. Figures 2.1 to 2.4 shows the estimated sizes of F_{nd} and F_d and the corresponding ADF tests for different sample sizes. The deviation from the nominal 5% size level decrease with the increasing sample size.

Figure 2.1 Size of statistic F_{nd} for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.



It is clear from the figures that the size of the tests is distorted when the value of δ_1 is close to -1 or 1. The reason is that the stationarity assumption of $\triangle y_t$ is violated for $|\delta_1| \geq 1$. All four tests have poor size characteristics for values of δ_1 close to 1. Furthermore, F_{nd} and F_d are also oversized, albeit less strongly, when δ_1 is close to -1. On the other hand, the ADF tests are not

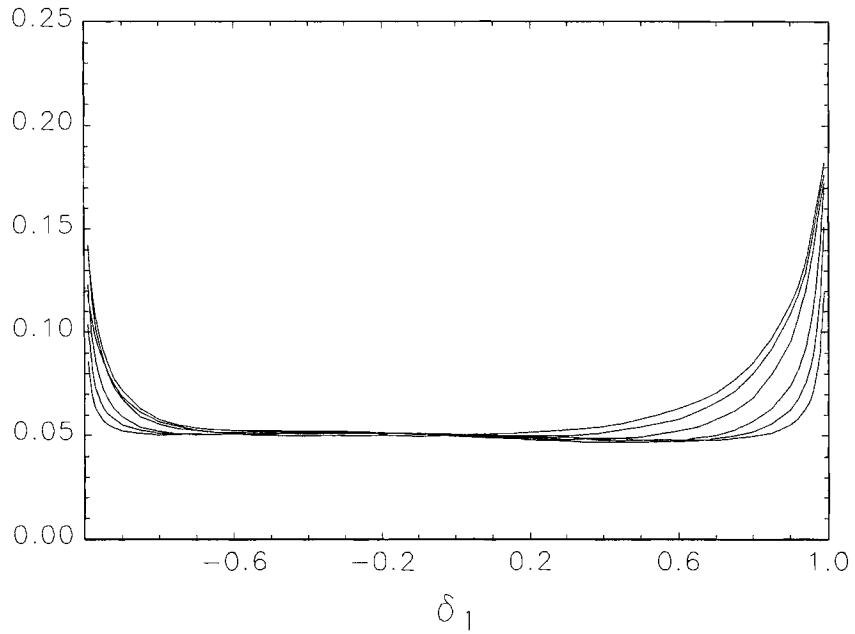
Figure 2.2 Size of statistic ADF_{nd} for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.



affected by small values of δ_1 in the same way. It is also worth noting that the two ADF tests are less distorted than their corresponding F tests, and that the two tests with a drift term, F_d and ADF_d, are not distorted as much as F_{nd} and ADF_{nd}. Furthermore, as may be expected, the deviation from the nominal 5% size level decreases with an increasing sample size. As the critical values in Table 2.1 are estimated for $\delta_1 = 0$, it is obvious that special attention is needed in practice if the sample size is small and the value of δ_1 is believed to be close to either -1 or 1.

In order to investigate how robust the tests are against non-normal errors, the empirical size of F_{nd} and F_d have been reestimated for errors drawn from the $t(6)$ and the $\chi^2(1) - 1$ distribution. The simulation results indicate that both F_{nd} and F_d are affected by the non-normal errors. For $t(6)$-distributed errors both F_{nd} and F_d show, for $T = 25$, about 1 to 2 percentage points higher size distortion at all values of δ_1 than the normal case. As the sample size increases the difference in size between the non-normal and normal size decreases. The size distortion is considerably larger for $\chi^2(1) - 1$ errors than it is for the $t(6)$-case. When $T = 25$, F_{nd} has at all values of δ_1 about 5 to 6 percentage points higher estimated size than it has for normal errors. At the

Figure 2.3 Size of statistic F_d for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.

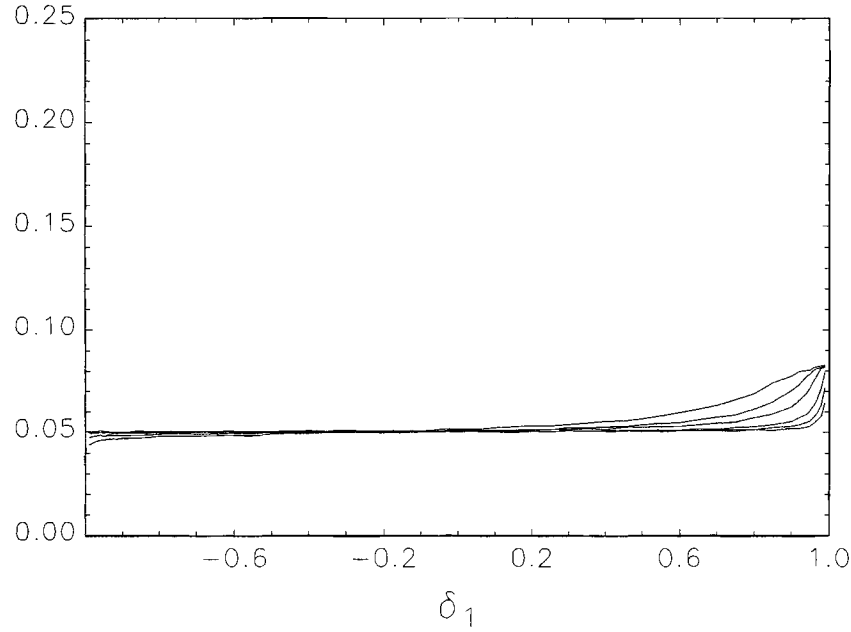


same time F_d shows a slightly smaller increase in size, about 3 to 4 percentage points. This diference diminishes with increasing sample size and becomes negligible when $T = 500$.

2.4.2 Bootstrapping the p-values

As the size is distorted for values of δ_1 close to 1 for all four tests, and also close to -1 for F_{nd} and F_d, an appropriate method of adjusting the size would be desirable. One way to prevent size distortion is to calculate new critical values for some particular value of δ_1, call it $\widehat{\delta}_1$. Another method would be to base the inference on bootstrap distributions of the tests. Size distortion can be diminished by obtaining the p-values by a bootstrap. Small sample properties of the tests are then considerably improved; see the survey by Li and Maddala (1996) for more information and details on bootstrapping time series, and Caner and Hansen (2001) who, when testing the unit root against the TAR model, based the inference on a bootstrap approximation to the asymptotic null distribution of the test statistic. The same bootstrap method was also used by Eklund (2003) when testing the unit root hypothesis against the STAR

Figure 2.4 Size of statistic ADF_d for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.



model. Using the bootstrap method to estimate p-values in stationary models requires that the test statistic is pivotal. Since the analytical limit expressions of F_{nd} and F_d are not given, it is not clear whether or not the test statistics are pivotal. However, results by Li and Maddala (1996) indicate that considering pivotal statistics may not be as important in the context of unit root models as it is in stationary models.

From Figures 2.1 and 2.3 it has already been possible to see how δ_1 influences the size of F_{nd} and F_d in small samples. Under H_{01}, the auxiliary model (2.4) simplifies to $y_t = \delta_1 \Delta y_{t-1} + y_{t-1} + \varepsilon_t$. Under H_{02}, the null model has the form $y_t = \delta_1 \Delta y_{t-1} + \alpha + y_{t-1} + \varepsilon_t$. The necessary p-values can then be obtained by a model-based bootstrap. This is also the case for the ADF tests. Consider first the resampling procedure for F_{nd}. Let $\widehat{\delta}_1$ and $\widehat{D}$ be the estimates of δ_1 and the distribution D of the error ε_t. Generate the bootstrap time series

$$y_t^b = \widehat{\delta}_1 \Delta y_{t-1}^b + y_{t-1}^b + \varepsilon_t^b, \; t = 1, \ldots, T, \tag{2.11}$$

where ε_t^b is a random draw from $\widehat{D}$, and the time series y_t^b, $t = 1, \ldots, T$, is the resampled bootstrap series. Initial values needed for resampling can be set to

sample values of the demeaned series y_t. The distribution of y_t^b is called the bootstrap distribution of the data. The value F_{nd}^b of F_{nd} is now obtained from the resampled series y_t^b. Repeating this B times yields B values of F_{nd} that constitute a realization from the distribution of F_{nd}, completely determined by $\widehat{\delta}_1$ and $\widehat{D}$. Defined by $p_T = P\left(F_{nd}^b > F_{nd}\right)$, the p-value p_T is in practice approximated by the frequency of the obtained F_{nd}^b that exceeds the observed value F_{nd}.

The resampling scheme can easily be modified to fit F_d. Including the estimate of the parameter α in the resampling model (2.11):

$$y_t^b = \widehat{\delta}_1 \Delta y_{t-1}^b + \widehat{\alpha} + y_{t-1}^b + \varepsilon_t^b,\ t = 1, \ldots, T. \tag{2.12}$$

The corresponding bootstrap distribution and the p-value p_T are then obtained in the same way as before.

2.4.3 Power simulations

In order to investigate the power properties of the tests, data is generated from the STAR model (2.1), i.e. the stationary alternative model. The fact that no analytical results are available for determining the parameter combinations yielding a stationary model constitutes a problem in a simulation study. By setting the transition function F to 0 or 1, some simple guidelines can be reached about stationarity, but no general conclusions can be drawn. Therefore, an approximative method is applied to determine when the model is nonstationary. By simulation, the model is taken to be nonstationary for a specific choice of parameters, if a realization exceeds a preset boundary with t. In this study, a realization of the alternative model, y_t, is said to originate from a nonstationary process if $|y_t| > \sigma t$ for $t > 1000000$ where σ equals the standard error of the errors ε_t in (2.1). This is of course just a rough indication on nonstationarity. For parameter choices on, or close to, the boundary between the stationary and the nonstationary regions, the approximation work less well, but probably good enough as a mean to compare the F_{nd} and F_d tests with the ADF tests.

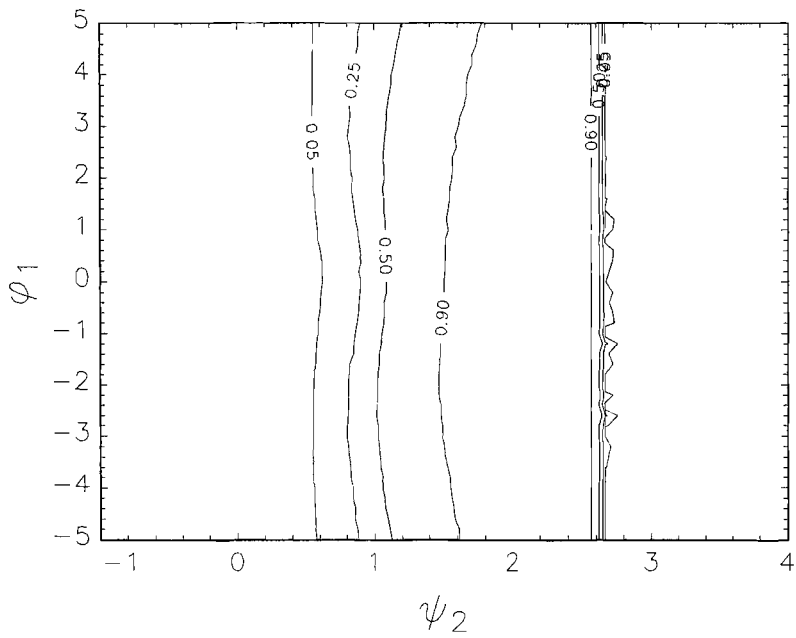
In the data generating process (2.1), $\theta_0 = 0$, $\varphi_0 = 0$, and the parameters in the nonlinear function $F(\cdot)$ are $\gamma = 10$, $c_1 = 1$ and $c_2 = -0.5$. For simplicity, $\theta_1 = -\varphi_1$, and in the same way, $\psi_1 = -\psi_2$. The alternative model then equals

$$\Delta y_t = -\varphi_1 \Delta y_{t-1} - \psi_2 y_{t-1} + (\varphi_1 \Delta y_{t-1} + \psi_2 y_{t-1})\, F(10, 1, -0.5, \Delta y_{t-1}) + \varepsilon_t, \tag{2.13}$$

where $\varepsilon_t \sim nid(0, 1)$. In (2.13), the magnitude of the regime shift is only determined by the two parameters φ_1 and ψ_2. Using a Monte Carlo simulation,

with 50 observations, 10000 replications, and 500 bootstrap replications for estimating the p-values, the power of the tests is estimated for a number of combinations of φ_1 and ψ_2. Contour plots of the power of F_{nd} and ADF_{nd} are depicted in Figures 2.5 and 2.6. Figure 2.7 shows for each combination of φ_1 and ψ_2 the difference in power between the tests, expressed as power of F_{nd} minus power of ADF_{nd}. Figures 2.8, 2.9, and 2.10 show the power and difference for the two other tests, F_d and ADF_d, in the same way.

Figure 2.5 Power of the F_{nd} statistic for $T = 50$ observations.

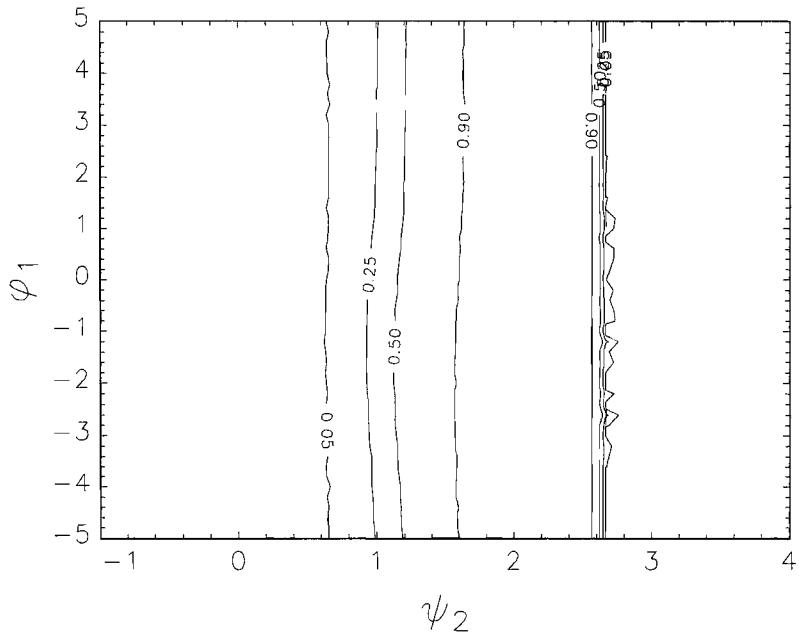


From Figures 2.5, 2.6, 2.8 and 2.9, it is seen that the STAR model is stationary in a vertical band of combinations of φ_1 and ψ_2. The value of ψ_2 range from about 0.5 to 2.7, whereas φ_1 does not seem to be restricted. It appears that it can take any value between -5 and 5. Outside this vertical band the model is nonstationary. All four tests have the strongest power for values of ψ_2 in the interval between about 1.6 and 2.6. But then, the gain from using F_{nd} or F_d instead of their Dickey Fuller counterparts is smallest in this specific interval, as seen in Figures 2.7 and 2.10. It is in fact negligible there and even slightly negative for some values of $\varphi_1 > 4$. Negative gain is also found for the F_d test for values of $\varphi_1 < -4$ observed in Figure 2.10. The strongest gains for both F tests are found for values of ψ_2 between 0.8 and 1.0. In this interval there are two separate regions with relatively large gains

in power.

The single largest gain for F_{nd} compared to ADF_{nd} is 18.9% percentage points more rejections of the true alternative hypothesis, the smallest is −5.6% percentage points. The corresponding largest and smallest gain for F_d are 16.6% and −6.4% percentage points. For about 10.6% of the combinations of φ_1 and ψ_2 the gain is negative for F_{nd}. The same figure for F_d is 12.7%.

Figure 2.6 Power of the ADF_{nd} statistic for $T = 50$ observations.



The main explanation why the gain is negative at some of the parameter combinations, is that the alternative STAR model is very close to the linear alternative model considered in the ADF test for these parameters. The auxiliary model (2.4) is then very close to or indistinguishable from the ADF model. This reduces the power of the tests compared to the ADF test because of the four extra parameters to be tested. As the area with positive gain dominates Figures 2.7 and 2.10 it appears safe to conclude that in general both F_{nd} and F_d have similar or higher power than the corresponding standard ADF tests when the alternative exhibit nonlinear behavior. Their use in situations where the STAR model is indeed an appropriate alternative can therefore be recommended.

Figure 2.7 Difference in power, $F_{nd} - ADF_{nd}$.

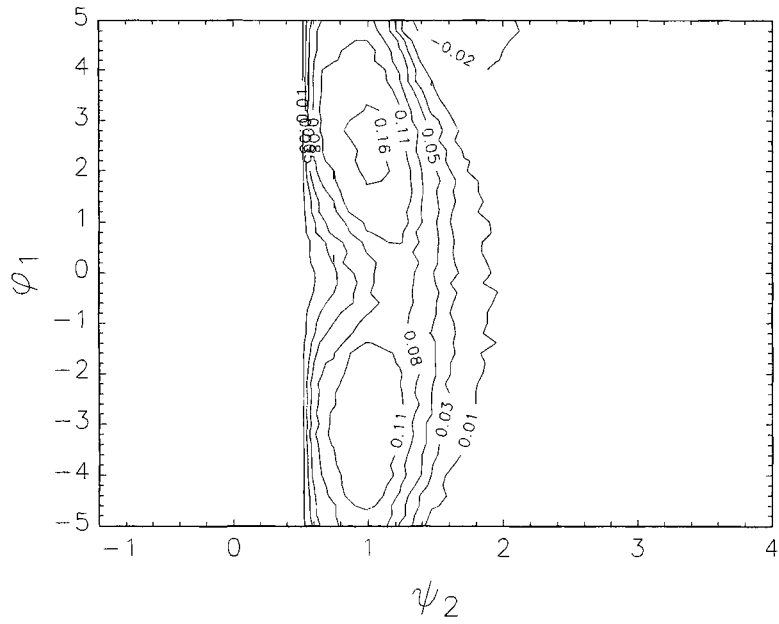


Figure 2.8 Power of the F_d statistic for $T = 50$ observations.

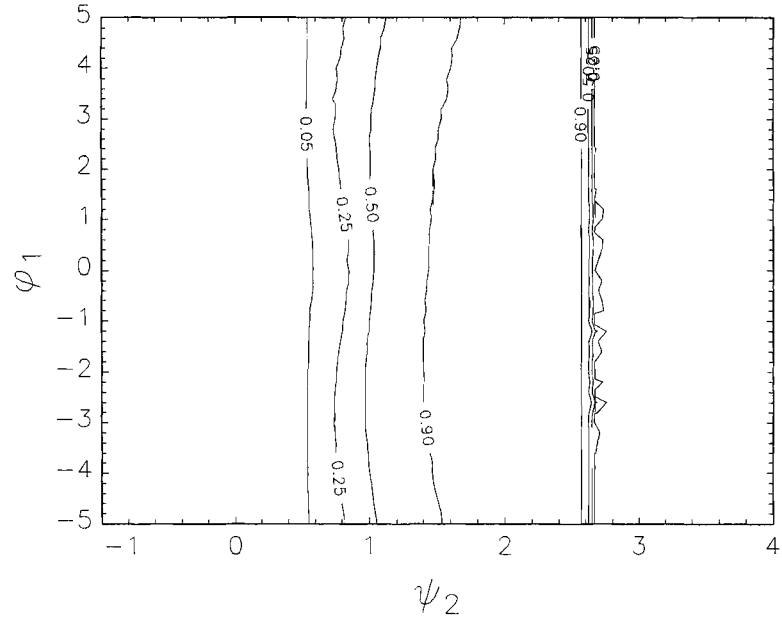


Figure 2.9 Power of the ADF_d statistic for $T = 50$ observations.

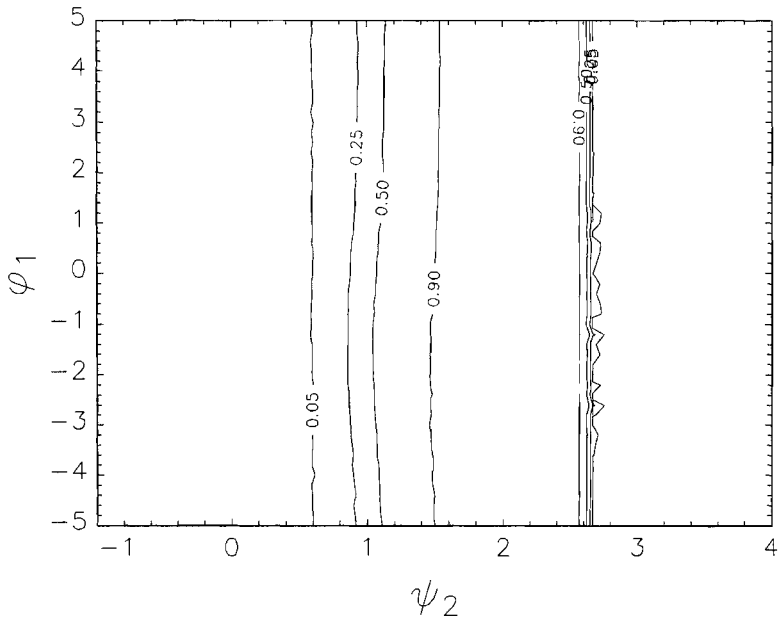
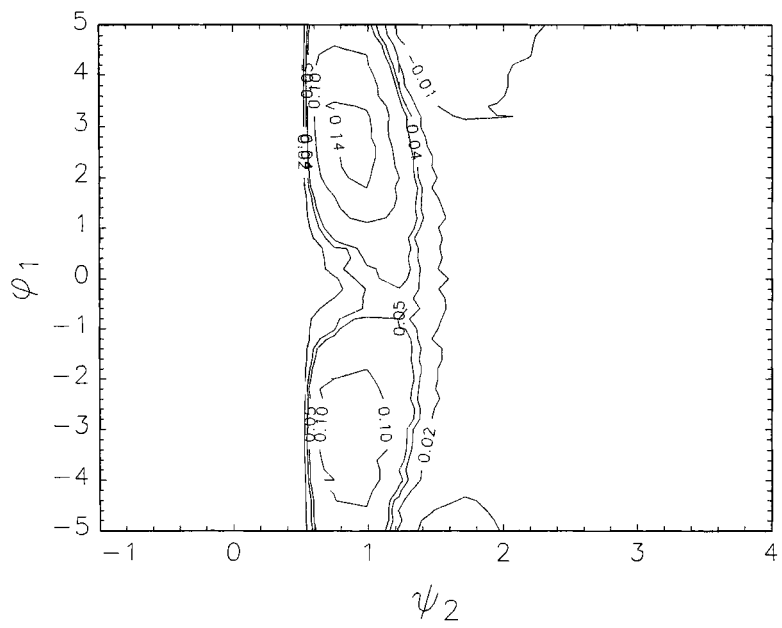


Figure 2.10 Difference in power, $F_d - ADF_d$.



2.5 Empirical application

The STAR model (2.1) is capable to characterizing asymmetric behavior in time series, and series with sudden upswings and downturns. When Δy_{t-1} is close to c_1 or c_2 in (2.2), the STAR model (2.1) behaves almost as if $\gamma = 0$, and possible as a unit root process depending on the values of θ_0 and ψ_1. On the other hand, if the difference between Δy_{t-1} and the parameter c_1, or c_2, is large, the model is nonlinear and stationary, implying a mean reverting behavior of the y_t process. These features of the second-order logistic STAR model make it attractive for modelling the real exchange rate for deviations above and below the equilibrium level. Determining the presence, or the absence, of a unit root in the real exchange rate is the main issue in testing the purchasing power parity (PPP) hypothesis. Before turning to the empirical application in this section, a brief discussion of the PPP literature will be given.

2.5.1 A short introduction to the PPP literature

In macroeconomic applications, and theory, the ability to be able to discriminate between stationary and nonstationary time series is of major importance. An area where this problem has received considerable attention is the one concerning the purchasing power parity (PPP) hypothesis, see Froot and Rogoff (1995) and Sarno and Taylor (2002) for two thorough surveys. The PPP corresponds to the idea that national price levels in different countries should tend to equal one another when expressed in a common currency. Price differences, or deviations from PPP, between two countries will be eliminated by arbitrage forces. As a consequence, for PPP to hold in the long-run, real exchange rates must be stationary. A theoretical insight into the deviations from the PPP, or long run equilibrium level, was given by Dumas (1992), who analyzed the dynamic process of the real exchange rate in spatially separated markets in the presence of proportional transactions costs. Dumas showed that deviations from PPP follow a nonlinear process that is mean reverting, and that the speed of adjustment varies directly on the magnitude of the deviation from PPP. This implies that the exchange rate will become increasingly mean reverting with the size of the deviation.

Stylized empirical findings indicate, on the other hand, high persistence in the deviations from PPP. O'Connell (1998) argued that it appears as if large deviations from PPP can be more persistent than small deviations, and that market frictions alone cannot account for the difficulty of detecting mean reversion in post-Bretton Woods real exchange rates. Lothian and Taylor (1996) argued that the high persistence of deviations from PPP, together with the

low power of standard unit root tests, may account for the widespread failure of empirical tests to support long-run PPP. By considering observations over a sample period of two centuries they where able to reject the unit root hypothesis in favor of a mean reverting process for two real exchange rates, U.S. dollar-pound sterling and French franc-pound sterling. Michael, Nobay and Peel (1997) rejected the linear framework in favor of an exponential smooth transition autoregressive (ESTAR) process, thus providing evidence of mean reverting behavior for PPP deviations. Taylor, Peel and Sarno (2001) considered a multivariate linear unit root test and provided empirical evidence that four major real bilateral dollar exchange rates are well characterized by nonlinear mean reverting processes, based on the ESTAR model.

Sarno and Taylor (2002) concluded that, at the present time, the long-run PPP seem to have some validity, at least for the major exchange rates, even though a number of problems have to be analyzed and resolved.

2.5.2 Testing the PPP hypothesis in practice

The real exchange rate for country i versus country j is expressed in logarithmic form as

$$z_t = p_{it} - s_t - p_{jt}, \quad t = 1, \ldots, T, \tag{2.14}$$

where s_t is the logarithm of the nominal exchange rate between country i and j expressed in country i's currency per country j's currency, and p_{it} and p_{jt} denote the logarithms of the consumer price index (CPI) for country i and j, respectively. The series z_t can then be interpreted as a measure of the deviation from the long-run steady state or PPP.

The real exchange rates are constructed for sixteen countries from the CPI series and the exchange rates defined as the price of US dollars in the currency of each home country. The data, consumer price indices and nominal exchange rates, are obtained from EcoWin and covers the following sixteen countries; Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, United Kingdom and USA. The sample consists of monthly observations from January 1960 to October 2002, except for five countries for which some early observations are missing. The sample for Germany only includes observations from 1968:1 to 2002:10, Japan from 1970:1 to 2002:10, the Netherlands from 1960:4 to 2002:10, Spain from 1961:1 to 2002:10, and, finally, Switzerland from 1974:1 to 2002:10. In total, 120 real exchange rates are constructed. For pairs of countries that both joined the common currency in January 1999, data only

up to December 1998 is used. Otherwise the largest available sample size is used for all pairs, which means that the length of the series varies between 344 and 512.

Tables 2.2 and 2.3 show the estimated p-values of F_{nd} and F_d based on 10000 replications. P-values less than 0.05 are printed in boldface. The results in Table 2.2 indicate that the random walk without drift can be rejected for 31 out of the 120 real exchange rates considered. When including a drift term under the null, the number of rejections increase to 44, as seen in Table 2.3. The corresponding number of rejections for ADF_{nd} and ADF_d are 7 and 23, respectively. The exchange rates for which the unit root is rejected can thus be considered stationary, giving support to the PPP hypothesis.

The test results both support and disagree with earlier work. Results in line with earlier studies, as in Bec et al. (2002), show that the PPP hypothesis can be supported for the real exchange rates BEL/GER, FIN/GER, FRA/GER, ITA/GER and UK/GER, and as in Taylor et al. (2001), for USA/GER. On the other hand, Bec et al. (2002) rejected the unit root hypothesis for another six exchange rates, Taylor et al. (2001) for another three. Conversely, for five of the significant real exchange rates considered in this paper, Bec et al. (2002) were unable to reject the unit root: CAN/BEL, USA/BEL, UK/FIN, USA/GER, and GER/CAN. They consider a smaller sample size, only from 1973:9 to 2000:9, which could explain some of the differences with this paper. Otherwise, as a whole, the proportion of rejected exchange rates is almost the same in this paper as in Bec et al. (2002).

The results in this paper can be viewed as a complement to earlier studies. Under the assumption that the PPP hypothesis holds, the two tests presented here, F_{nd} and F_d, have low power discriminating a random walk from a stationary nonlinear process at sample sizes available for the tests. The same fact holds for most univariate unit root tests. However, Taylor et al. (2001) noted that, somewhat paradoxically, the failure to reject a unit root may indicate that the real exchange rate has, on average, been relatively close to equilibrium, rather than implying that no such long-run equilibrium exists. In their estimated ESTAR model, the real exchange rate will be closer to a unit root process the closer it is to its long-run equilibrium.

2.6 Conclusions

In this paper, two F-type tests are proposed for the joint unit root and linearity hypothesis against a second-order logistic smooth transition autoregressive (STAR) model. The tests allows one to discriminate between nonstationary

Table 2.2 Estimated p-values to the F_{nd} test. Bold values corresponds to a significant test at the 5% level.

vs	BEL	CAN	DEN	FIN	FRA	GER	ITA	JAP	NL	NOR	SPA	SWE	SWI	UK	USA
AUS	0.23	0.091	0.054	0.34	0.60	**0.0006**	0.089	0.81	0.32	0.76	0.26	0.29	**0.022**	**0.033**	**0.0021**
BEL	-	**0.0034**	**0.040**	0.44	0.51	**0.020**	0.13	0.76	0.34	0.63	0.20	0.69	**0.0068**	0.27	**0.0018**
CAN		-	0.072	0.37	0.087	0.10	0.46	0.49	0.10	0.093	0.42	0.45	0.12	0.42	0.13
DEN			-	0.23	0.70	**0.048**	0.19	0.63	0.051	**0.0058**	0.59	0.18	**0.030**	0.062	**0.0001**
FIN				-	0.59	0.069	0.43	0.31	0.41	0.15	**0.013**	0.48	**0.012**	**0.0058**	0.49
FRA					-	**0.044**	0.30	0.48	0.48	0.38	0.39	0.96	**0.0001**	0.41	0.13
GER						-	**0.039**	0.88	0.091	0.33	**0.038**	0.52	**0.0056**	**0.033**	**0.0046**
ITA							-	0.26	0.11	0.24	0.15	0.16	**0.0004**	0.10	0.63
JAP								-	0.85	0.69	0.23	0.17	0.77	0.41	0.70
NL									-	0.074	0.44	0.39	**0.0065**	**0.043**	**0.0018**
NOR										-	0.30	0.38	0.15	0.052	**0.0014**
SPA											-	**0.017**	**0.0089**	0.10	0.52
SWE												-	0.10	0.24	0.24
SWI													-	**0.012**	0.15
UK														-	0.23

Table 2.3 Estimated p-values to the F_d test. Bold values corresponds to a significant test at the 5% level.

vs	BEL	CAN	DEN	FIN	FRA	GER	ITA	JAP	NL	NOR	SPA	SWE	SWI	UK	USA
AUS	0.12	**0.036**	**0.015**	0.21	0.89	**0.0003**	**0.037**	0.72	0.17	0.81	0.31	0.17	**0.0051**	**0.0050**	**0.0004**
BEL	-	**0.0012**	**0.027**	0.32	0.46	**0.011**	0.065	0.74	0.25	0.50	0.11	0.87	**0.0004**	0.12	**0.0011**
CAN		-	**0.045**	0.30	0.066	**0.031**	0.35	0.41	0.064	0.063	0.31	0.36	**0.036**	0.32	0.058
DEN			-	0.14	0.65	**0.014**	0.069	0.61	**0.035**	**0.0031**	0.83	0.096	**0.0028**	**0.011**	**0.0002**
FIN				-	0.49	**0.028**	0.25	0.17	0.30	0.077	**0.0084**	0.38	**0.0052**	**0.0043**	0.40
FRA					-	**0.023**	0.15	0.43	0.38	0.27	0.27	0.93	**0.0000**	0.31	0.084
GER						-	**0.033**	0.73	**0.020**	0.18	**0.027**	0.85	**0.0019**	**0.0023**	**0.0004**
ITA							-	0.20	**0.036**	0.11	0.12	0.064	**0.0001**	0.059	0.78
JAP								-	0.87	0.68	0.13	0.073	0.65	0.32	0.71
NL									-	0.059	0.50	0.28	**0.0004**	**0.0086**	**0.0010**
NOR										-	0.19	0.24	0.059	**0.0092**	**0.0007**
SPA											-	**0.014**	**0.0046**	0.062	0.54
SWE												-	**0.030**	0.17	0.16
SWI													-	**0.0005**	0.067
UK														-	0.12

and stationary time series. This is important in statistical analysis and empirical applications. Some new limit results, extending earlier work, and critical values for the F-tests are presented. As the alternative model is well suited for modelling real exchange rates, the two tests are applied to a number of real exchange rates as an illustration. The test results complement earlier studies. Support to the purchasing power parity (PPP) hypothesis is at any rate provided for 44 out of 120 real exchange rates considered in this work.

Appendix A

Proof of Theorem 1

(*a*) Let u_t, defined in (2.6), satisfy Assumption 1 and let $\sum_{j=0}^{\infty} |\omega_j| < \infty$. Set $X_t = u_{t-1}^2 \varepsilon_t$. X_t is then a martingale difference sequence. Furthermore,

$$\begin{aligned} EX_t^2 &= \sigma^2 E u_t^4 \\ E|X_t|^{2+r} &= E\left|u_{t-1}^2 \varepsilon_t\right|^{2+r} = E|u_t|^{4+2r} E|\varepsilon_t|^{2+r} < \infty. \end{aligned} \tag{A.1}$$

so that $\{X_t\}$ is a uniformly integrable sequence. Now,

$$T^{-1}\sum_{t=1}^{T} X_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^4 \varepsilon_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^4 \left(\varepsilon_t^2 - \sigma^2\right) + T^{-1}\sigma^2 \sum_{t=1}^{T} u_{t-1}^4. \tag{A.2}$$

Set $Z_t = u_{t-1}^4 \left(\varepsilon_t^2 - \sigma^2\right)$. It follows that $\{Z_t\}$ is a martingale difference sequence. Furthermore,

$$E|Z_t|^{1+r} = E\left|u_{t-1}^4 \left(\varepsilon_t^2 - \sigma^2\right)\right|^{1+r} = E|u_t|^{4+4r} E\left|\varepsilon_t^2 - \sigma^2\right|^{1+r} < \infty \tag{A.3}$$

implies that

$$T^{-1}\sum_{t=1}^{T} Z_t = T^{-1}\sum_{t=1}^{T} u_{t-1}^4 \left(\varepsilon_t^2 - \sigma^2\right) \xrightarrow{p} 0. \tag{A.4}$$

Also,

$$T^{-1}\sigma^2 \sum_{t=1}^{T} u_{t-1}^4 \xrightarrow{p} \sigma^2 E u_t^4 \tag{A.5}$$

which gives the result

$$T^{-1/2}\sum_{t=1}^{T} u_{t-1}^2\varepsilon_t \Rightarrow \sigma\sqrt{Eu_t^4}W(1). \tag{A.6}$$

(*b*) Let u_t and ε_t be as in (*a*). Set $X_t = u_{t-1}^3\varepsilon_t$. Then $\{X_t\}$ is a martingale difference sequence. Furthermore,

$$\begin{aligned} EX_t^2 &= \sigma^2 Eu_t^6 \\ E|X_t|^{2+r} &= E\left|u_{t-1}^3\varepsilon_t\right|^{2+r} = E|u_t|^{6+3r}E|\varepsilon_t|^{2+r} < \infty \end{aligned} \tag{A.7}$$

so that $\{X_t\}$ is uniformly integrable. Now,

$$T^{-1}\sum_{t=1}^{T} X_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^6\varepsilon_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^6\left(\varepsilon_t^2-\sigma^2\right)+T^{-1}\sigma^2\sum_{t=1}^{T} u_{t-1}^6. \tag{A.8}$$

Set $Z_t = u_{t-1}^6\left(\varepsilon_t^2-\sigma^2\right)$. Then $\{Z_t\}$ is a martingale difference sequence. For $r > 0$,

$$E|Z_t|^{1+r} = E\left|u_{t-1}^6\left(\varepsilon_t^2-\sigma^2\right)\right|^{1+r} = E|u_t|^{6+6r}E\left|\varepsilon_t^2-\sigma^2\right|^{1+r} < \infty \tag{A.9}$$

which implies

$$T^{-1}\sum_{t=1}^{T} Z_t = T^{-1}\sum_{t=1}^{T} u_{t-1}^6\left(\varepsilon_t^2-\sigma^2\right) \xrightarrow{p} 0. \tag{A.10}$$

Furthermore,

$$T^{-1}\sigma^2\sum_{t=1}^{T} u_{t-1}^6 \xrightarrow{p} \sigma^2 Eu_t^6 \tag{A.11}$$

which gives the desired result

$$T^{-1/2}\sum_{t=1}^{T} u_{t-1}^3\varepsilon_t \Rightarrow \sigma\sqrt{Eu_t^6}W(1). \tag{A.12}$$

(*c*) Let $v_t = \left(u_t, u_{t-1}^2\varepsilon_t\right)'$, $V_t = \sum_{i=1}^{t} v_i$ and $V_0 = 0$. Then, since $T^{-1/2}V_T \Longrightarrow \left(\lambda W(1), \sigma\sqrt{Eu_t^4}W(1)\right)'$, Hansen (1992), Theorem 4.1, states that the

elements of the sum

$$T^{-1}\sum_{t=1}^{T} V_{t-1}v_t' = T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \xi_{t-1} \\ \sum_{i=1}^{t-1} u_{i-1}^2\varepsilon_i \end{bmatrix}\begin{bmatrix} u_t & u_{t-1}^2\varepsilon_t \end{bmatrix} \quad \text{(A.13)}$$

$$= T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \xi_{t-1}u_t & \xi_{t-1}u_{t-1}^2\varepsilon_t \\ u_t\sum_{i=1}^{t-1} u_{i-1}^2\varepsilon_i & u_{t-1}^2\varepsilon_t\sum_{i=1}^{t-1} u_{i-1}\varepsilon_i \end{bmatrix}$$

will converge weakly to some stochastic integrals. In particular,

$$T^{-1}\sum_{t=1}^{T}\xi_{t-1}u_{t-1}^2\varepsilon_t \Rightarrow \sigma\sqrt{Eu_t^4}\lambda\int_0^1 W(r)\,dB(r) + \Lambda_{1,2} \quad \text{(A.14)}$$

where $\Lambda_{1,2}$ is element $(1,2)$ in the matrix

$$\Lambda = \lim_{T\longrightarrow\infty} T^{-1}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} Ev_iv_j' = \quad \text{(A.15)}$$

$$= \lim_{T\longrightarrow\infty} T^{-1}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\begin{bmatrix} u_i \\ u_{i-1}^2\varepsilon_i \end{bmatrix}\begin{bmatrix} u_j & u_{j-1}^2\varepsilon_j \end{bmatrix}$$

$$= \lim_{T\longrightarrow\infty} T^{-1}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\begin{bmatrix} u_iu_j & u_iu_{j-1}^2\varepsilon_j \\ u_ju_{i-1}^2\varepsilon_i & u_{i-1}^2u_{j-1}^2\varepsilon_j\varepsilon_i \end{bmatrix}$$

Then, since $u_iu_{j-1}^2$ and ε_j are independent for $j \geq i+1$,

$$\Lambda_{1,2} = \lim_{T\longrightarrow\infty} T^{-1}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\left(u_iu_{j-1}^2\varepsilon_j\right) \quad \text{(A.16)}$$

$$= \lim_{T\longrightarrow\infty} T^{-1}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\left(u_iu_{j-1}^2\right)E\left(\varepsilon_j\right) = 0.$$

This implies that the two Brownian motions $W(r)$ and $B(r)$ are independent, and the result follows.

(*d*) Consider

$$T^{-1}\sum_{t=1}^{T}\xi_tu_t^4 = T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^4 - Eu_t^4\right) + T^{-1}\sum_{t=1}^{T}\xi_tEu_t^4. \quad \text{(A.17)}$$

Now, since $\xi_t = \xi_{t-1} - u_t$, the first sum on the right-hand side of (A.17) equals

$$T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^4 - Eu_t^4\right) = T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^4 - Eu_t^4\right) + T^{-1}\sum_{t=1}^{T}u_t\left(u_t^4 - Eu_t^4\right) \tag{A.18}$$

where the last term is $O_p(1)$. Now let $v_t = \left(u_t, u_t^4 - Eu_t^4\right)'$, $V_t = \sum_{i=1}^{t} v_i$ and $V_0 = 0$. Then, from Hansen (1992), Theorem 4.1, it follows that the sum $T^{-1}\sum_{t=1}^{T} V_{t-1}v_t'$ converges weakly to a stochastic integral. Therefore, as a consequence, $T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^4 - Eu_t^4\right) = O_p(1)$ and

$$T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^4 = T^{-3/2}E\left(u_t^4\right)\sum_{t=1}^{T}\xi_t + o_p(1) \Rightarrow Eu_t^4\lambda\int_0^1 W(r)\,dr. \tag{A.19}$$

(*e*) First consider

$$T^{-1}\sum_{t=1}^{T}\xi_t u_t^5 = T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^5 - Eu_t^5\right) + T^{-1}\sum_{t=1}^{T}\xi_t Eu_t^5. \tag{A.20}$$

Now, since $\xi_t = \xi_{t-1} - u_t$, the first sum on the right-hand side of (A.20) equals

$$T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^5 - Eu_t^5\right) = T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^5 - Eu_t^5\right) + T^{-1}\sum_{t=1}^{T}u_t\left(u_t^5 - Eu_t^5\right) \tag{A.21}$$

where the last term is $O_p(1)$ as before. Now let $v_t = \left(u_t, u_t^5 - Eu_t^5\right)'$, $V_t = \sum_{i=1}^{t} v_i$ and $V_0 = 0$. Then, again using Hansen (1992), Theorem 4.1, one can conclude that the sum $T^{-1}\sum_{t=1}^{T} V_{t-1}v_t'$ converges weakly to a stochastic integral. Thus, $T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^5 - Eu_t^5\right) = O_p(1)$ and

$$T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^5 = T^{-3/2}E\left(u_t^5\right)\sum_{t=1}^{T}\xi_t + o_p(1) \Rightarrow Eu_t^5\lambda\int_0^1 W(r)\,dr. \tag{A.22}$$

(*f*) As a starting-point, consider the sum

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^3 = T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^3 - Eu_t^3\right) + T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2Eu_t^3 \tag{A.23}$$

and let $v_t = \left(u_t, u_t^3 - Eu_t^3\right)'$, V_t and V_0 be as before. Let Assumption 1 hold with $\eta = 3$. It then follows from Hansen (1992), Theorem 4.2, that the sum $T^{-3/2}\sum_{t=1}^{T}\left(V_{t-1}\otimes V_{t-1}\right)v_t'$ converges weakly to a stochastic integral. This implies that

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^3 - Eu_t^3\right) = O_p\left(1\right) \tag{A.24}$$

and

$$T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^3 = T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2Eu_t^3 + o_p\left(1\right) \Rightarrow E\left(u_t^3\right)\lambda^2\int_0^1 W^2\left(r\right)dr. \tag{A.25}$$

What remains to show is that $\underset{T\longrightarrow\infty}{plim}\, T^{-2}\sum_{t=1}^{T}\xi_t^2u_t^3 = \underset{T\longrightarrow\infty}{plim}\, T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^3$. It is easily shown that

$$\xi_t^2 - \xi_{t-1}^2 = 2\xi_{t-1}u_t + u_t^2. \tag{A.26}$$

The difference between the two sums in (A.25) is given by

$$\sum_{t=1}^{T}\xi_t^2u_t^3 - \sum_{t=1}^{T}\xi_{t-1}^2u_t^3 = \sum_{t=1}^{T}\left(\xi_t^2 - \xi_{t-1}^2\right)u_t^3 = 2\sum_{t=1}^{T}\xi_{t-1}u_t^4 + \sum_{t=1}^{T}u_t^5, \tag{A.27}$$

where the first sum on the right-hand side is $O_p\left(T^{3/2}\right)$ from (*d*) above, and the second sum is $O_p\left(T\right)$. This implies that $T^{-2}\sum_{t=1}^{T}\left(\xi_t^2 - \xi_{t-1}^2\right)u_t^3 = o_p\left(T^2\right)$, and

$$T^{-2}\sum_{t=1}^{T}\xi_t^2u_t^3 \Rightarrow E\left(u_t^3\right)\lambda^2\int_0^1 W^2\left(r\right)dr \tag{A.28}$$

as desired.

(g) Begin by considering the sum

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^4 = T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^4 - Eu_t^4\right) + T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2Eu_t^4 \tag{A.29}$$

and let $v_t = \left(u_t, u_t^4 - Eu_t^4\right)'$, V_t and V_0 be as before. Let Assumption 1 hold with $\eta = 3$. It then follows from Hansen (1992), Theorem 4.2, that the sum $T^{-3/2}\sum_{t=1}^{T}\left(V_{t-1}\otimes V_{t-1}\right)v_t'$ converges weakly to a stochastic integral. This implies that

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^4 - Eu_t^4\right) = O_p\left(1\right) \tag{A.30}$$

and

$$T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^4 = T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2Eu_t^4 + o_p\left(1\right) \Rightarrow Eu_t^4\lambda^2\int_0^1 W^2\left(r\right)dr. \tag{A.31}$$

Using the same idea as in the proof of (f),

$$\sum_{t=1}^{T}\left(\xi_t^2 - \xi_{t-1}^2\right)u_t^4 = 2\sum_{t=1}^{T}\xi_{t-1}u_t^5 + \sum_{t=1}^{T}u_t^6, \tag{A.32}$$

where the first sum on the right-hand side is $O_p\left(T^{3/2}\right)$ from (e) above, and the second one is $O_p\left(T\right)$. The result then follows since

$$T^{-2}\sum_{t=1}^{T}\xi_t^2u_t^4 = T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2u_t^4 + o_p\left(T\right) \Rightarrow Eu_t^4\lambda^2\int_0^1 W^2\left(r\right)dr. \tag{A.33}$$

This concludes the proof of Theorem 1. ■

Bibliography

[1] Bec, F., Salem, M. B., and Carrasco, M., (2002), "The purchasing power parity puzzle: Evidence from a LSTAR specification", Working Paper, Université de Paris I.

[2] Berben, R.-P., and van Dijk, D., (1999), "Unit root tests and asymmetric adjustment: A reassessment", Working paper, Erasmus University Rotterdam.

[3] Caner, M., and Hansen, B. E., (2001), "Threshold autoregression with a unit root", *Econometrica*, 69, 1555-1596.

[4] Dumas, B., (1992), "Dynamic equilibrium and the real exchange rate in a spatially separated world", *Review of Financial Studies*, 5, 153-180.

[5] Eklund, B., (2003), "Testing the unit root hypothesis against the logistic smooth transition autoregressive model", Working Paper, Stockholm School of Economics.

[6] Enders, W., and Granger, C. W. J., (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.

[7] Froot, K. A., and Rogoff, K., (1995), "Perspectives on PPP and long-run real exchange rates", in G. Grossman and K. Rogoff (eds.), *Handbook of International Economics*, North-Holland, Amsterdam, 1647-1688.

[8] Granger, C. W. J., and Teräsvirta, T., (1993), *Modelling nonlinear economic relationships*, Oxford, Oxford University Press.

[9] Hamori, S., and Tokihisa, A., (1997), "Testing for a unit root in the presence of a variance shift", *Economics Letters*, 57, 245-253.

[10] Hansen, B. E., (1992), "Convergence to stochastic integrals for dependent heterogeneous processes", *Econometric Theory*, 8, 489-500.

[11] Kapetanios, G., and Shin, Y. (2000), "Testing for a linear unit root against nonlinear threshold stationarity", Discussion paper, http://www.econ.ed.ac.uk/Academic_00Discuss.htm, University of Edingburgh.

[12] Kapetanios, G., Shin, Y., and Snell, A., (2003), "Testing for a unit root in the nonlinear STAR framework", *Journal of Econometrics*, 112, 359-379.

[13] Leybourne, S. J., Mills, T. C., and Newbold, P., (1998), "Spurious rejections by Dickey-Fuller tests in the presence of a break under the null", *Journal of Econometrics*, 87, 191-203.

[14] Li, H., and Maddala, G. S., (1996), "Bootstrapping time series models", *Econometric Reviews*, 15, 115-158.

[15] Lothian, J. R., and Taylor, M. P., (1996), "Real exchange rate behavior: The recent float from the perspective of the past two centuries, *Journal of Political Economy*, 104, 488-509.

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

[17] Michael, P., Nobay, R. A., and Peel, D. A., (1997), "Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation", *Journal of Political Economy*, 105, 862-879.

[18] Nelsen, C. R., Piger, J., and Zivot, E., (2001), "Markov regime switching and unit root tests", *Journal of Business and Economic Statistics*, 19, 404-415.

[19] O'Connell, P. G. J., (1998), "Market frictions and real exchange rates", *Journal of International Money and Finance*, 17, 71-95.

[20] Perron, P., (1989), "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 57, 1361-1401.

[21] Pippenger, M. K., and Goering, G. E., (1993), "A note on the empirical power of unit root tests under threshold processes", *Oxford Bulletin of Economics and Statistics*, 55, 473-481.

[22] Sarno, L., and Taylor, M. P., (2002), "Purchasing power parity and the real exchange rate", IMF Staff Papers, 49, No. 1.

[23] Taylor, M. P., Peel, D. A., and Sarno, L., (2001), "Nonlinear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles", *International Economic Review*, 42, 1015-1042.

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

[25] Teräsvirta, T., (1994b), "Testing linearity and modelling nonlinear time series", *Kybernetika*, 30, 319-330.

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

Chapter 3

Testing the constancy of the error covariance matrix in vector models

3.1 Introduction

Estimating the parameters of an econometric model is necessary for any use of the model, be it forecasting or policy evaluation. Finding out thereafter whether or not the model appears to satisfy the assumptions under which it was estimated should be an integral part of a normal modelling exercise. An evaluation through misspecification tests has long been standard practice in univariate time series analysis, see Box and Jenkins (1970). It has been advocated for multivariate regression models as well, and a number of misspecification tests exist for linear and nonlinear single-equation models.

Misspecification tests also exist for estimated vector models such as the linear vector autoregressive (VAR) model, although their use in the VAR context seems much less widespread than in the single-equation case. The available tests are generally intended for testing the specification of the conditional mean. It is argued, however, that sometimes misspecification of the conditional mean shows in the error variance, making it time-varying. Besides, the error variance of a model may be nonconstant even when the conditional mean is correctly specified. It would therefore be desirable to have statistical tools for investigating possible parameter nonconstancy in vector models.

In this paper we consider the problem of testing the constancy of the error covariance matrix of a vector model. Many approaches to testing the constancy of the error variance in single-equation models exist in the literature, whereas the same is not true for constancy of the error covariance matrix in a multivariate system. A special case of the test discussed in this paper may be viewed as a multivariate generalization of the heteroskedasticity test of White (1980), and another special case generalizes the test against autoregressive conditional heteroskedasticity of Engle (1982). Yet another variant of the test generalizes the constant variance test of Medeiros and Veiga (2003) who assumed that under the alternative, the error variance changes smoothly over time. An important assumption in the present paper is that while the error variances change over time, the correlations between them remain constant. This restriction has the considerable advantage that it decreases the dimension of the null hypothesis compared to the case where both variances and covariances can fluctuate freely under the alternative hypothesis.

The plan of the paper is as follows. The model and the structure of the error covariance matrix are introduced in Section 2 as well as the null hypothesis to be tested. The test statistic is presented in Section 3 and illustrated in the bivariate case in Section 4. Section 5 contains the test against smoothly changing variances. Small sample properties of the test are considered by simulation in Section 6. Section 7 contains conclusions.

3.2 The model

Consider the following multivariate (vector) model

$$\mathbf{y}_t = \mathbf{f}(\mathbf{x}_t; \boldsymbol{\psi}) + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{3.1}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$ is an $(m \times 1)$ vector, $\mathbf{x}_t = (x_{1t}, \ldots, x_{nt})'$ is an $(n \times 1)$ vector of explanatory variables, some of which may be lags of the elements of $\mathbf{y}_t$, $\boldsymbol{\psi} = (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2', \ldots, \boldsymbol{\psi}_m')'$ is a column vector of model parameters, and

$$\mathbf{f}(\mathbf{x}_t; \boldsymbol{\psi}) = (f(\mathbf{x}_t; \boldsymbol{\psi}_1), \ldots, f(\mathbf{x}_t; \boldsymbol{\psi}_m))'$$

is an $(m \times 1)$ vector of linear or nonlinear functions of $\mathbf{x}_t$. The m-dimensional vector error process $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of random variables with zero mean and conditional $(m \times m)$ covariance matrix $\boldsymbol{\Omega}_t | I_t = [\omega_{ijt}]$, where I_t is the information set at time t. More specifically, the time-varying variances ω_{iit} have the form

$$\omega_{iit} = h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it}), \quad i = 1, \ldots, m, \tag{3.2}$$

where $h_i(\cdot)$ is a positive function of a linear combination of a $((p_i + 1) \times 1)$ parameter vector $\boldsymbol{\varphi}_i = \left(\sigma_i^2, \varphi_{i1}^2, \varphi_{i2}^2, \ldots, \varphi_{ip_i}^2\right)'$ and a vector of stochastic or deterministic variables $\mathbf{v}_{it} = (1, v_{i1t}, v_{i2t}, \ldots, v_{ip_it})'$ believed to influence the variance. In some situations $\mathbf{v}_{it} = \mathbf{v}_t$, $i = 1, \ldots, m$. The variance ω_{iit} is thus a function of a time-varying component. We make the following general assumption:

Assumption 1 *The time-varying covariances ω_{ijt} have the form*

$$\omega_{ijt} = \rho_{ij} (\omega_{iit} \omega_{jjt})^{1/2}; \quad i, j = 1, \ldots, m, \tag{3.3}$$

where ρ_{ij} is the correlation coefficient of ε_{it} and ε_{jt}.

Bollerslev (1990) made this assumption when he defined the Constant Conditional Correlation GARCH (CCC-GARCH) model. It makes the correlation structure of $\boldsymbol{\Omega}_t$ time-invariant and implies that the covariances are time-varying functions only through the corresponding variances. This is a simplifying assumption which imposes structure on the way the variances are allowed to change over time. It can be tested as well, see Tse (2000). Both Engle (2002) and Tse and Tsui (2002) consider generalizations to the CCC-GARCH model. In this work, however, the focus is on testing the constancy of the whole covariance matrix when the alternative is characterized by (3.2) and (3.3). The general functions $h_i(\cdot)$ defined in (3.2) allow for a wide variety

of covariance structures. Using the above notation, the null hypothesis to be tested is

$$H_0 : \omega_{iit} = \omega_{ii} \text{ or } \boldsymbol{\varphi}_i' \mathbf{v}_{it} = \sigma_i^2, \quad i = 1, \ldots, m, \tag{3.4}$$

since under the null hypothesis $\omega_{iit} = h_i\left(\sigma_i^2\right)$, $i = 1, \ldots, m$, that is, the variances remain constant over time. Under H_0, the errors are independent.

The general definition (3.2) allows us to test the constancy of the variances against a number of alternatives. The flexibility implied in choosing functions h_i implicitly reduces the severity of the assumption of constant conditional correlations because h_i represents a whole set of positive-valued functions. The only, albeit nontrivial, restriction is that the argument of h_i is a linear combination of elements of $\mathbf{v}_{it}$.

An interesting specification of the variance, considered by White (1980) as an alternative to homoskedasticity, is obtained by defining

$$\omega_{iit} = h_i\left(\sigma_i^2 + \sum_{j=1}^{n}\sum_{k=j}^{n} \delta_{ijk} x_{jt} x_{kt}\right) = h_i\left(\sigma_i^2 + \boldsymbol{\delta}_i' vech\left(\mathbf{x}_t \mathbf{x}_t^{'}\right)\right), \quad i = 1, \ldots, m, \tag{3.5}$$

where $\boldsymbol{\delta}_i = (\delta_{i11}, \delta_{i12}, \ldots, \delta_{i1n}, \delta_{i22}, \delta_{i23}, \ldots, \delta_{i2n}, \ldots, \delta_{inn})'$, $i = 1, \ldots, m$, and $\mathbf{x}_t$ is defined above. The hypothesis of constant variance is equivalent to $H_0 : \delta_{ijk} = 0$, $j = 1, \ldots, n$, $k = j, \ldots, n$, $i = 1, \ldots, m$, or $\boldsymbol{\delta}_i = 0$, $i = 1, \ldots, m$. Another interesting case is the one of conditional heteroskedasticity:

$$\omega_{iit} = h_i\left(\sigma_i^2 + \sum_{j=1}^{q} \alpha_{ij} \varepsilon_{i,t-j}^2\right), \quad \alpha_{ij} \geq 0, \quad i = 1, \ldots, m, \tag{3.6}$$

where the null hypothesis is $H_0 : \alpha_{ij} = 0$, $j = 1, \ldots, q$, $i = 1, \ldots, m$. If h_i is assumed to be an identity function, the alternative to a time-invariant covariance matrix is seen to be the Constant Conditional Correlation ARCH model of order q, see Cecchetti, Cumby and Figlewski (1988) and Bollerslev (1990). Additional examples will be considered in Section 5.

3.3 The test statistic

In this section we derive the Lagrange Multiplier test for the null hypothesis (3.4) under standard regularity conditions. Parameterizing the linear combination $\boldsymbol{\varphi}_i' \mathbf{v}_{it}$ in the variance equation (3.2) in a suitable fashion yields the appropriate test statistics in the cases of interest. In order to derive the test statistic we need the log-likelihood function of model (3.1). It follows from Assumption 1 that the covariance matrix $\boldsymbol{\Omega}_t$ can be written in the form

$$\boldsymbol{\Omega}_t = \mathbf{D}_t \mathbf{P} \mathbf{D}_t \tag{3.7}$$

where

$$\begin{aligned}\mathbf{D}_t &= \operatorname{diag}\left[\omega_{11t}^{1/2},\ldots,\omega_{mmt}^{1/2}\right] \\ &= \operatorname{diag}\left[\left[h_1\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)\right]^{1/2},\ldots,\left[h_m\left(\boldsymbol{\varphi}_m'\mathbf{v}_{mt}\right)\right]^{1/2}\right] \end{aligned} \tag{3.8}$$

is a diagonal matrix of standard deviations, and $\mathbf{P} = \left[\rho_{ij}\right]$ is the correlation matrix. The parameters $\boldsymbol{\theta}$ in the covariance matrix are ordered according to $\boldsymbol{\theta} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \ldots, \boldsymbol{\varphi}_m', \boldsymbol{\rho}')'$, where $\boldsymbol{\varphi}_i$ is the column vector of parameters in ω_{iit} from (3.2), and $\boldsymbol{\rho}$ is the column vector of correlations ordered as follows:

$$\boldsymbol{\rho} = \left(\rho_{12}, \rho_{13}, \ldots, \rho_{1m}, \rho_{23}, \rho_{24}, \ldots, \rho_{2m}, \ldots, \rho_{(m-1)m}\right)'. \tag{3.9}$$

As defined above $\boldsymbol{\varphi}_i$ is a $((p_i + 1) \times 1)$ vector, and let $p_\rho = m(m-1)/2$ be the number of parameters in $\boldsymbol{\rho}$. Then $p = p_1 + \ldots + p_m + m + p_\rho$ is the number of parameters in $\boldsymbol{\theta}$. We make an additional assumption which is not overly restrictive:

Assumption 2 *The functions $f(\mathbf{x}_t; \boldsymbol{\psi}_i)$ and $h_i(\boldsymbol{\varphi}_i'\mathbf{v}_t)$, $i = 1, \ldots, m$, are at least twice continuously differentiable for $\boldsymbol{\psi}_i$ and $\boldsymbol{\varphi}_i$ (parameter space) almost everywhere in the sample space and, furthermore, $\boldsymbol{\psi}_i$ and $\boldsymbol{\varphi}_i$ are variation-free.*

The assumption of $\boldsymbol{\psi}_i$ and $\boldsymbol{\varphi}_i$ being variation-free makes the population information matrix block diagonal in $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$, and in the subsequent considerations we can ignore the part of the score vector related to $\boldsymbol{\psi}$. The quasi-log likelihood function for observation t has the form

$$\begin{aligned}\ell_t(\boldsymbol{\psi}, \boldsymbol{\theta}) &= c + \frac{1}{2}\ln\left|\boldsymbol{\Omega}_t^{-1}\right| - \frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\varepsilon}_t \\ &= c + \ln\left|\mathbf{D}_t^{-1}\right| - \frac{1}{2}\ln|\mathbf{P}| - \frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\varepsilon}_t. \end{aligned} \tag{3.10}$$

The relevant block of the average score vector and the population information matrix are defined in the following lemma:

Lemma 1 *The 'variance block' of the average score vector of the quasi log-likelihood function (3.10) equals*

$$\begin{aligned}\bar{\mathbf{s}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_1^T \Bigg\{ &\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{D}_t - \frac{1}{2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \frac{1}{2}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) \\ &+ \frac{1}{2}\frac{\partial vec(\mathbf{P})'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \mathbf{P}^{-1}\right)\Bigg\}, \end{aligned} \tag{3.11}$$

where $\frac{\partial vec(\mathbf{D}_t^{-1})'}{\partial \boldsymbol{\theta}}$ *and* $\frac{\partial vec(\mathbf{P})'}{\partial \boldsymbol{\theta}}$ *are* $(p \times m^2)$ *matrices, and the two vectorized matrices are* $(m^2 \times 1)$*. The corresponding* $(p \times p)$ *block of the population information matrix equals*

$$\begin{aligned}
\mathbf{I}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_1^T E\Bigg\{ & \frac{\partial vec(\mathbf{D}_t^{-1})'}{\partial \boldsymbol{\theta}} \{\mathbf{D}_t \otimes \mathbf{D}_t + \\
& + \frac{1}{2}(\mathbf{P}^{-1} \otimes \boldsymbol{\Omega}_t) + \frac{1}{2}(\boldsymbol{\Omega}_t \otimes \mathbf{P}^{-1})\Bigg\} \frac{\partial vec(\mathbf{D}_t^{-1})}{\partial \boldsymbol{\theta}'} \\
& - \frac{1}{2}\frac{\partial vec(\mathbf{D}_t^{-1})'}{\partial \boldsymbol{\theta}}(\mathbf{D}_t \otimes \mathbf{P}^{-1} + \mathbf{P}^{-1} \otimes \mathbf{D}_t)\frac{\partial vec(\mathbf{P})}{\partial \boldsymbol{\theta}'} \\
& - \frac{1}{2}\frac{\partial vec(\mathbf{P})'}{\partial \boldsymbol{\theta}}(\mathbf{D}_t \otimes \mathbf{P}^{-1} + \mathbf{P}^{-1} \otimes \mathbf{D}_t)\frac{\partial vec(\mathbf{D}_t^{-1})}{\partial \boldsymbol{\theta}'} \\
& + \frac{1}{2}\frac{\partial vec(\mathbf{P})'}{\partial \boldsymbol{\theta}}(\mathbf{P}^{-1} \otimes \mathbf{P}^{-1})\frac{\partial vec(\mathbf{P})}{\partial \boldsymbol{\theta}'}\Bigg\}.
\end{aligned} \tag{3.12}$$

Proof. See the appendix.

The relevant blocks of the average score vector and the population information matrix can thus be expressed in terms of the matrices $\mathbf{D}_t$, $\mathbf{P}$, $\boldsymbol{\Omega}_t$ and the first order partial derivatives of $\mathbf{D}_t^{-1}$ and $\mathbf{P}$ with respect to the parameter vector $\boldsymbol{\theta}$. In the Appendix, an expression for $\frac{\partial vec(\mathbf{D}_t^{-1})}{\partial \boldsymbol{\theta}'}$ can be found, together with three explicit examples for the bivariate case. Estimating the parameters $\boldsymbol{\theta}$ under the null hypothesis, estimates of the average score vector, $\bar{\mathbf{s}}_T\left(\widetilde{\boldsymbol{\theta}}\right)$, and the population information matrix, $\widetilde{\mathbf{I}}_T\left(\widetilde{\boldsymbol{\theta}}\right)$, are given by replacing $\frac{\partial vec(\mathbf{D}_t^{-1})}{\partial \boldsymbol{\theta}'}$, $\frac{\partial vec(\mathbf{P})}{\partial \boldsymbol{\theta}'}$, $\mathbf{D}_t$, $\mathbf{P}$, $\boldsymbol{\Omega}_t$ and $\boldsymbol{\varepsilon}_t$ by their maximum likelihood estimators under H_0. These are denoted by $\left.\frac{\partial vec(\mathbf{D}_t^{-1})}{\partial \boldsymbol{\theta}'}\right|_{H_0}$, $\left.\frac{\partial vec(\mathbf{P})}{\partial \boldsymbol{\theta}'}\right|_{H_0}$, $\widetilde{\mathbf{D}}$, $\widetilde{\mathbf{P}}$, $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{f}\left(\mathbf{x}_t; \widetilde{\boldsymbol{\psi}}\right)$, respectively. The subscript t in $\mathbf{D}_t$ and $\boldsymbol{\Omega}_t$ can be dropped under the null, since the variances are constant. Under standard regularity conditions, the LM test statistic given by

$$LM = T\,\bar{\mathbf{s}}_T\left(\widetilde{\boldsymbol{\theta}}\right)' \widetilde{\mathbf{I}}_T\left(\widetilde{\boldsymbol{\theta}}\right)^{-1} \bar{\mathbf{s}}_T\left(\widetilde{\boldsymbol{\theta}}\right) \tag{3.13}$$

is asymptotically χ^2 distributed, with $p - p_\rho - m = p_1 + \ldots + p_m$ degrees of freedom under the null hypothesis (3.4), where $\widetilde{\mathbf{I}}_T\left(\widetilde{\boldsymbol{\theta}}\right)$ is a consistent estimator of $\mathbf{I}_T(\boldsymbol{\theta})$ under H_0. The number of degrees of freedom equals the number of parameter restrictions in H_0.

3.4 Bivariate illustration

In this section we consider a number of examples of the test. In order to illustrate its structure, we first discuss the bivariate case in which the alternative only contains a single time-varying covariance. In this case, average score vector in Lemma 1 equals

$$\bar{\mathbf{s}}_T(\boldsymbol{\theta}) = \frac{1}{2T(1-\rho^2)} \sum_1^T (\mathbf{G}_t \mathbf{k}_t + \mathbf{r}_t) \tag{3.14}$$

where ρ is the correlation between the two error terms $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$, and $\mathbf{k}_t = (k_{1t}, 0, 0, k_{2t})'$, with $k_{it} = \phi_{it}^2 - \rho\phi_{1t}\phi_{2t} - (1-\rho^2)$, and $\phi_{it} = \varepsilon_{it}/\omega_{iit}^{1/2}$, $i = 1, 2$. Furthermore, $\mathbf{G}_t$ is a $(p \times 4)$ matrix defined as

$$\mathbf{G}_t = \begin{bmatrix} \mathbf{g}_{1t} & \mathbf{0}_1 & \mathbf{0}_1 & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{g}_{2t} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\omega_{11t}} \frac{\partial \omega_{11t}}{\partial \boldsymbol{\varphi}_1} & \mathbf{0}_1 & \mathbf{0}_1 & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \frac{1}{\omega_{22t}} \frac{\partial \omega_{22t}}{\partial \boldsymbol{\varphi}_2} \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.15}$$

where $\mathbf{g}_{it}$ is the weighted partial derivatives of the variance ω_{iit} with respect to the parameter vector $\boldsymbol{\varphi}_i$ in ω_{iit}, and $\mathbf{0}_i$ is a $((p_i + 1) \times 1)$ null column vector, $i = 1, 2$. The first $p_1 + p_2 + 2$ rows in $\mathbf{G}_t$ correspond to the first $p_1 + p_2 + 2$ parameters in $\boldsymbol{\theta}$, while the last row corresponds to the correlation coefficient ρ. Finally, the column vector $\mathbf{r}_t$ in (3.14) equals

$$\mathbf{r}_t = \frac{-\rho\phi_{1t}^2 - \rho\phi_{2t}^2 + (1+\rho^2)\phi_{1t}\phi_{2t} + \rho(1-\rho^2)}{(1-\rho^2)} [\mathbf{0}_1', \mathbf{0}_2', 2]'. \tag{3.16}$$

Replacing ρ, ϕ_{it}, ω_{iit}, and ε_{it} by their maximum likelihood estimators $\widetilde{\rho}$, $\widetilde{\phi}_{it}$, $\widetilde{\sigma}_i^2$, and $\widetilde{\varepsilon}_{it}$ respectively, the average score evaluated under H_0 becomes

$$\bar{\mathbf{s}}_T\left(\widetilde{\boldsymbol{\theta}}\right) = \frac{1}{2T(1-\widetilde{\rho}^2)} \sum_1^T \widetilde{\mathbf{G}}_t \widetilde{\mathbf{k}}_t \tag{3.17}$$

since $\frac{1}{T}\sum_1^T \widetilde{\mathbf{r}}_t = 0$. The expression for the population information matrix is given by

$$\mathbf{I}_T(\boldsymbol{\theta}) = \frac{1}{2T(1-\rho^2)} \sum_1^T E\left\{\mathbf{G}_t \mathbf{A}_t \mathbf{G}_t' - \mathbf{G}_t \mathbf{B}_t' - \mathbf{B}_t \mathbf{G}_t' + \mathbf{C}\right\} \tag{3.18}$$

where ρ and $\mathbf{G}_t$ are defined as before, and, letting $\mathbf{0}$ be a $((p_1+p_2+2)\times 1)$ null column vector,

$$\mathbf{A}_t = \begin{bmatrix} \frac{2-\rho^2}{2\omega_{11t}^2} & 0 & 0 & -\frac{\rho^2}{2\omega_{11t}\omega_{22t}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{\rho^2}{2\omega_{11t}\omega_{22t}} & 0 & 0 & \frac{2-\rho^2}{2\omega_{22t}^2} \end{bmatrix}$$
$$\mathbf{B}_t = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{\rho}{\omega_{11t}} & 0 & 0 & \frac{\rho}{\omega_{22t}} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{0}\mathbf{0}' & \mathbf{0} \\ \mathbf{0}' & \frac{2(1+\rho^2)}{1-\rho^2} \end{bmatrix}. \tag{3.19}$$

Evaluated under the null hypothesis, $\widetilde{\mathbf{A}}_t = \widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{B}}_t = \widetilde{\mathbf{B}}$, and the population information matrix becomes

$$\widetilde{\mathbf{I}}_T\left(\widetilde{\boldsymbol{\theta}}\right) = \frac{1}{2\left(1-\widetilde{\rho}^2\right)T}\sum_1^T \left\{\widetilde{\mathbf{G}}_t\widetilde{\mathbf{A}}\widetilde{\mathbf{G}}_t' - \widetilde{\mathbf{G}}_t\widetilde{\mathbf{B}}' - \widetilde{\mathbf{B}}\widetilde{\mathbf{G}}_t' + \widetilde{\mathbf{C}}\right\}. \tag{3.20}$$

Consider now the first one of the two variance specifications mentioned in Section 2, that is,

$$h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right) = h_i\left(\sigma_i^2 + \boldsymbol{\delta}_i' vech\left(\mathbf{x}_t\mathbf{x}_t'\right)\right), \; i = 1,2.$$

The vectors $\mathbf{g}_{it}$ are then defined as

$$\begin{aligned} \mathbf{g}_{it} &= \frac{1}{\omega_{iit}}\frac{\partial \omega_{iit}}{\partial \boldsymbol{\varphi}_i} = \frac{1}{\omega_{iit}}\left[\begin{array}{cc} \frac{\partial h_i}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}\frac{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}{\partial \sigma_i^2} & \frac{\partial h_i}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}\frac{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}{\partial \boldsymbol{\delta}_i'} \end{array}\right]' \\ &= \frac{1}{\omega_{iit}}\frac{\partial h_i}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}\left[\begin{array}{cc} 1, & vech\left(\mathbf{x}_t\mathbf{x}_t'\right)' \end{array}\right]', \; i = 1,2. \end{aligned} \tag{3.21}$$

Note that $p_1 = p_2 = n(n+1)/2$ so that the null vector blocks in $\mathbf{B}$ and $\mathbf{C}$ are $((2+n(n+1))\times 1)$ column vectors. Note, furthermore, that under H_0, $\frac{\partial h_i}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}} = c_i$ and, as discussed in Breusch and Pagan (1979) the LM test statistic will be independent of the actual functional form of the function h_i. As the null hypothesis equals $H_0 : \boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \mathbf{0}$, the resulting LM test statistic in (3.13) has an asymptotic $\chi^2(n(n+1))$ distribution under the null hypothesis.

For the second case with conditional heteroskedasticity,

$$h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right) = h_i\left(\sigma_i^2 + \sum_{j=1}^q \alpha_{ij}\varepsilon_{i,t-j}^2\right), \quad \alpha_{ij} \geq 0, \quad i = 1,2, \tag{3.22}$$

the matrix $\mathbf{G}_t$ follows in the same way, with $\mathbf{g}_{it} = \frac{1}{\omega_{iit}}\frac{\partial h_i}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}\left(1, \varepsilon_{it-1}^2, \ldots, \varepsilon_{it-q}^2\right)'$, $i = 1,2$. The null vector blocks in $\mathbf{B}_t$ and $\mathbf{C}$ are now of size $(2(q+1)\times 1)$.

Under the null hypothesis $H_0 : \alpha_{ij} = 0$, $i = 1, 2$, $j = 1, \ldots, q$, the LM test statistic has an asymptotic $\chi^2(2q)$ distribution.

Remark 1 *Note that the above explicit calculations of the average score vector and the population information matrix are just given as examples. Since the score vector and the information matrix in Lemma 1 are expressed in terms of the matrices* $\mathbf{D}_t$, $\mathbf{P}$, $\mathbf{\Omega}_t$ *as well as the first-order partial derivatives of* $\mathbf{D}_t^{-1}$ *and* $\mathbf{P}$, *an operational test statistic is obtained simply by replacing these matrices in Lemma 1 directly by their maximum likelihood estimators under the null hypothesis.*

3.5 Testing against smoothly changing variances

3.5.1 Assumptions

The test can easily be modified to also allow for variances that change smoothly over time. We will only consider testing the hypothesis of constant variance against variances with two extreme regimes, where the argument of h_i changes monotonically over time. It is straightforward to generalize the method into situations where the change of the argument is non-monotonic. To accomplish a modified test, the time-varying variances ω_{iit}, equation (3.2), is respecified according to

$$\omega_{iit} = h_i\left(\sigma_i^2 + \lambda_i F_i(z_{it})\right), \quad i = 1, \ldots, m, \tag{3.23}$$

where λ_i is a scalar parameter and $F_i(\cdot)$ is a real-valued function of a time varying component z_{it}. Assume, for notational simplicity, that $F_i(0) = 0$, $i = 1, \ldots, m$. This is not a restrictive assumption as we can always replace $F_i(z_{it})$ by $\widetilde{F}_i(z_{it}) = F_i(z_{it}) - F_i(0)$. To proceed further, we make the following two assumptions:

Assumption 3 *Function* $F_i(z)$ *is bounded, odd, monotonically increasing, possessing a nonzero derivative of order* $(2d+1)$ *in an open interval* $(-a, a)$ *around* $z = 0$ *for* $a > 0$. *Furthermore,* $\partial^h F_i(z)/\partial z^h\big|_{z=0} \neq 0$ *for* h *odd and* $1 \leq h \leq 2d+1$.

Assumption 4 $\sigma_i^2 + \lambda_i F_{i,\min} > 0$ *and* $\sigma_i^2 + \lambda_i F_{i,\max} > 0$, $i = 1, \ldots, m$, *where* $F_{i,\min} = F(-\infty)$ *and* $F_{i,\max} = F(\infty)$, $i = 1, \ldots, m$.

The two assumptions implies a smooth change in the arguments between $\sigma_i^2 + \lambda_i F_{i,\min}$ and $\sigma_i^2 + \lambda_i F_{i,\max}$. When h_i, $i = 1, \ldots, m$, are identity functions, then the variances fluctuate between these values. Furthermore, when

$F_i(z_{it}) = 0$ for all t, the variances are constant. Using the above notation the null hypothesis to be tested is

$$H_0 : \omega_{iit} = \omega_{ii} \text{ or } F_i(z_{it}) = 0, \quad i = 1, \ldots, m. \tag{3.24}$$

The alternative hypothesis is simply that (3.24) is not valid.

3.5.2 The identification problem and a solution

The testing problem outlined above is nonstandard in the sense that the variance functions (3.23) are only identified under the alternative. When the null hypothesis holds, λ_i is an unidentified nuisance parameter. Hansen (1996) contains a thorough discussion of this situation and considers remedies. In this paper we follow Luukkonen et al. (1988) and circumvent the problem by a suitable approximation to the log-likelihood function, more specifically the diagonal elements of $\mathbf{D}_t$. This is done by replacing $F_i(z_{it})$ by its Taylor expansion around $z_{it} = 0$, the value under the null hypothesis. Suppose that Assumption 3 holds for $d = 0$ and consider the first-order Taylor expansion $w_i z_{it} + R_i$ where $w_i = \partial F_i / \partial z_{it}|_{z_{it}=0}$ and R_i is the remainder. The function can then be approximated by

$$F_i(z_{it}) = w_i z_{it} + R_i(z_{it}). \tag{3.25}$$

Inserting the approximation (3.25) into (3.23) and reparameterizing yields $\omega_{iit} = h_i\left(\sigma_i^2 + \lambda_i^* z_{it} + \lambda_i R_i\right)$, where $\lambda_i^* = \lambda_i w_i$. The i-th component of the null hypothesis is then $\lambda_i^* = 0$. Note that we can ignore the remainder because $R_i \equiv 0$, $i = 1, \ldots, m$, under the null hypothesis. Thus, we can write

$$\begin{aligned} \mathbf{D}_t^* &= \operatorname{diag}\left[\left[h_1\left(\sigma_1^2 + \lambda_1^* z_{1t}\right)\right]^{1/2}, \ldots, \left[h_m\left(\sigma_m^2 + \lambda_m^* z_{mt}\right)\right]^{1/2}\right] \\ &= \operatorname{diag}\left[\left[h_1\left(\boldsymbol{\varphi}_1' \mathbf{v}_{1t}^*\right)\right]^{1/2}, \ldots, \left[h_m\left(\boldsymbol{\varphi}_m' \mathbf{v}_{mt}^*\right)\right]^{1/2}\right] \end{aligned} \tag{3.26}$$

and use (3.26) to approximate $\mathbf{D}_t$ in (3.10), where $\boldsymbol{\varphi}_i = \left(\sigma_i^2, \lambda_i^*\right)'$ and $\mathbf{v}_{it}^* = (1, z_{1t})'$. This matrix has a structure similar to (3.8) in that the argument of h_i is a linear combination of terms such that h_i is constant when the null hypothesis holds. Thus, after reparameterizing, (3.26) and (3.8) result in the same log-likelihood function. As a consequence, Lemma 1 is also valid for testing the hypothesis of constant variances against time-varying variances with regime shifts.

3.5.3 Smooth and deterministically time-varying variances

As a leading example, consider the logistic function

$$F_i(z_{it}) = (1 + \exp(-z_{it}))^{-1} - 1/2, \quad i = 1, \ldots, m. \tag{3.27}$$

The transition function $F(\cdot)$ defines a smooth transition of the variances over time between two regimes corresponding to the cases when $F(\cdot) = -1/2$ and $F(\cdot) = 1/2$. A natural choice is to define the time-varying component as

$$z_{it} = \gamma_i (\tau_{it} - c_i), \; i = 1, \ldots, m, \tag{3.28}$$

where τ_{it} is the transition variable, the parameter γ_i controls the speed of adjustment between the regimes, $\gamma_i > 0$ for identification purposes, and c_i is the transition centre. Function (3.27) satisfies Assumption 3, and assuming $\sigma_i^2 \pm \lambda_i/2 > 0$ for $i = 1, \ldots, m$, even Assumption 4 is satisfied.

The null hypothesis $\mathbf{\Omega}_t \equiv \mathbf{\Omega}$ equals $H_0 : \gamma_i = 0$, $i = 1, \ldots, m$, whereas at least one $\gamma_i > 0$ under the alternative. Under the null hypothesis $F(\cdot) = 0$, and the variances remains constant over time, $\omega_{iit} = \sigma_i^2$, $i = 1, \ldots, m$. It should be noted that letting $\gamma_i \to \infty$ in (3.28) creates a break in the variance ω_{iit} at $\tau_{it} = c_i$, since the logistic function then equals a step function.

In testing parameter constancy, an obvious choice is to let $\tau_{it} = t/T$ in (3.27), see Lin and Teräsvirta (1994). In this case the variances are changing smoothly and deterministically over time. Note that the transition in variance is defined in terms of its relative location so that τ_{it} is bounded between 0 and 1. This assumption is made for convenience. If the null model is rejected and the alternative estimated, parameter c_i has a straightforward interpretation as an indicator of the relative mid-point of the change in the variance ω_{iit}. Explicit expressions of the LM test statistic can easily be calculated as in the previous section.

Consider as before the bivariate case where the variances under the alternative are approximated by $\omega_{iit} = h_i\left(\sigma_i^2 + \lambda_i^* t/T\right)$, $i = 1, 2$, and the null hypothesis, $H_0 : \lambda_1^* = \lambda_2^* = 0$. Defining the matrix $\mathbf{G}_t$ as in (3.15), now with $\mathbf{g}_{it} = \frac{1}{\omega_{iit}} \frac{\partial h_i}{\partial \boldsymbol{\varphi}_i' \mathbf{v}_{it}} \begin{pmatrix} 1, & t/T \end{pmatrix}'$, $i = 1, 2$, the average score vector evaluated under the null equals

$$\bar{\mathbf{s}}_T\left(\widetilde{\boldsymbol{\theta}}\right) = \frac{1}{2T\left(1 - \widetilde{\rho}^2\right)} \sum_1^T \widetilde{\mathbf{G}}_t \widetilde{\mathbf{k}}_t \tag{3.29}$$

where $\widetilde{\mathbf{k}}_t = (k_{1t}, 0, 0, k_{2t})'$ is defined as before. The population information matrix under the null is given by

$$\widetilde{\mathbf{I}}_T\left(\widetilde{\boldsymbol{\theta}}\right) = \frac{1}{2T\left(1 - \widetilde{\rho}^2\right)} \sum_1^T \left\{ \widetilde{\mathbf{G}}_t \widetilde{\mathbf{A}} \widetilde{\mathbf{G}}_t' - \widetilde{\mathbf{G}}_t \widetilde{\mathbf{B}}' - \widetilde{\mathbf{B}} \widetilde{\mathbf{G}}_t' + \widetilde{\mathbf{C}} \right\} \tag{3.30}$$

where the coefficient matrices $\mathbf{A}_t$, $\mathbf{B}_t$, and $\mathbf{C}$ are defined as before in (3.19). The resulting LM test statistic can then be expressed explicitly as

$$\begin{aligned} LM &= T\,\bar{\mathbf{s}}_T\left(\widetilde{\boldsymbol{\theta}}\right)'\widetilde{\mathbf{I}}_T\left(\widetilde{\boldsymbol{\theta}}\right)^{-1}\bar{\mathbf{s}}_T\left(\widetilde{\boldsymbol{\theta}}\right) \\ &= \frac{3}{T^3\left(1-\widetilde{\rho}^2\right)^2}\left(\left(2-\widetilde{\rho}^2\right)\widetilde{\sigma}_1^4\left(\sum_1^T \widetilde{k}_{1t}t\right)^2\right. \\ &\quad \left. +2\widetilde{\rho}^2\widetilde{\sigma}_1^2\widetilde{\sigma}_2^2\sum_1^T \widetilde{k}_{1t}t\sum_1^T \widetilde{k}_{2t}t + \left(2-\widetilde{\rho}^2\right)\widetilde{\sigma}_2^4\left(\sum_1^T \widetilde{k}_{2t}t\right)^2\right). \quad (3.31) \end{aligned}$$

Since H_0 has two restrictions the LM statistic has an asymptotic $\chi^2(2)$ distribution under the null hypothesis.

Note that in the present case the arguments of h_i are assumed to change monotonically over time under the alternative, $\tau_{it} = t/T$. Replacing this variance structure by a higher-order polynomial in t/T leads to more flexible tests. The only adjustment one has to make for such a modification is to recalculate the column vectors $\mathbf{g}_{it}$, $i = 1,\ldots,m$, and adjust the degrees of freedom. Finally, for $m = 1$ the test statistic (3.31) collapses into the one presented in Medeiros and Veiga (2003) for testing the constancy of a single error variance over time.

3.6 Small sample properties of the test

In order to investigate the empirical size of the test in small samples, and obtain an idea of the power the test has against the alternatives considered, we conduct a Monte Carlo simulation for the case $m = 2$ in (3.1). The simulations are performed at a 5% level of significance for sample sizes $25, 50, 100, 250, 500, 1000$ and with 100000 replications in each experiment. The size and power are estimated for the joint test but are also conducted separately for each equation.

3.6.1 Size simulations

When considering the size properties, we assume $h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right) = \boldsymbol{\varphi}_i'\mathbf{v}_{it}$ in (3.2) for simplicity because, as Breusch and Pagan (1979) point out, the test is independent of the functional form of h_i. The vector error process $\{\boldsymbol{\varepsilon}_t\}$ in (3.1) is assumed bivariate normal with mean zero and covariance matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad (3.32)$$

where $\rho = -0.9, 0.0$ and 0.9. Note that the elements of ε_t become linearly dependent and $\mathbf{\Omega}$ singular when $\rho = \pm 1$. The log-likelihood function is thus not well defined if $\rho = \pm 1$, which may cause size distortion in cases where $|\rho| \approx 1$.

In testing against the ARCH(q), we let the lag length $q = 1, 3, 5$, and in the White case the number of explanatory variables in the model (3.1) is assumed to equal 2 or 4. In this case, the explanatory variables x_{jt}, $j = 1, \ldots, n$, are assumed independent normal with mean $1/j$ and variance 1.

Tables 3.1 to 3.3 show the estimated rejection frequencies when testing against the three alternatives for the different combinations at a 5% level of significance. No major size distortion is detected in any of the experiments. Nor does $|\rho| = 0.9$ lead to any overrejection of the null hypothesis compared to the case $\rho = 0$. The estimated size tends towards the nominal level with increasing sample size, as expected.

The test against the White specification, Table 3.1, is slightly oversized for $n = 4$. When $n = 2$, the distortion is smaller. The joint test has a slightly higher rejection frequency than the two univariate tests.

Table 3.2 contains results for testing against ARCH(1) errors. In this case the test is undersized for small samples. No difference can be detected between the univariate and joint tests. The results for ARCH(3) and ARCH(5) errors are similar and not reported here.

The results in Table 3.3 indicate that the test against smoothly changing variances with $\tau_{it} = t/T$ has very good size properties. The test is only slightly undersized at small samples and attains about the correct size at $T = 250$.

3.6.2 Violating Assumption 1: time-varying correlations

Before considering the power of the test, it is important to investigate the consequences of violating Assumption 1. The assumption of constant correlations can be criticized as being too restrictive. It may be argued that if the null hypothesis of constant variances is rejected, it would be difficult without any further investigation to distinguish between a rejection due to time-varying variances with constant correlation, time-varying correlations or a combination of the two. In order to investigate the consequences of time-varying correlations, we set up a Monte Carlo simulation similar to the one used in the size simulations. Generating data under the null hypothesis, errors assumed bivariate normal with constant variances, but with time-varying correlation, the effects of this alternative on the size of the test can be analyzed. We investigate two different types of time-varying correlations. First we assume that the correlation grows monotonically from -1 to 1 over the sample

period $t = 1, \ldots, T$. In the remaining case, the correlation is drawn randomly for each observation from a uniform distribution. The two time-varying correlation structures can be summarized as follows:

$$\begin{array}{ll} 1. & \rho_t = -1 + \frac{2(t-1)}{T-1} \\ 2. & \rho_t \sim U(-1, 1) \end{array}$$

where $t = 1, \ldots, T$. In order to save space, not all combinations of variance specifications and time-varying correlations are reported. The tables contain only a selection of the simulation results. The omitted results are available from the authors upon request.

Table 3.4 contains the estimated rejection frequencies for the different setups with time-varying correlation. They can be compared directly with the results in Tables 3.1-3.2. No major difference in rejection frequencies suggests that the test does not have appreciable power against time-varying correlations.

It can be seen for the White case with four explanatory variables that the rejection frequencies are slightly higher than the corresponding values for the size simulations in Table 3.1. For $n = 2$ the difference is smaller than with $n = 4$. These differences are in practice negligible, which means that any effect of time-varying correlation can be ignored. The same conclusion holds for the ARCH(1) case, in which no difference is found in the rejection frequencies compared to the size simulations. The same is true for ARCH(3) and ARCH(5) errors. Simulating the test against the smooth transition alternative with a random correlation, type 2, likewise shows no power against time-varying correlation. On the other hand, letting the correlation increase monotonically over time, results in strong underrejection of the null hypothesis. This is a somewhat surprising result considering the strong structure of the sequence of correlations. A proper explanation can not be given at this point. It is well-known that changes in variance affect the amplitude of errors, introducing time-dependence in the volatility. While leaving the amplitude intact, it could be possible that changes in correlation introduces some sort of 'co-volatility' of the variances, affecting the size of the test.

3.6.3 Power simulations

In power simulations we again assume that $h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it}) = \boldsymbol{\varphi}_i' \mathbf{v}_{it}$. The test is carried out against the White specification with $n = 2, 4$, the ARCH(q) specification for $q = 1, 3, 5$, and furthermore, the smoothly changing variance with $\tau_{it} = t/T$ as before. The vector error process $\{\boldsymbol{\varepsilon}_t\}$ is drawn from a bivariate

normal distribution with mean zero but now with the covariance matrix

$$\mathbf{\Omega}_t = \begin{bmatrix} \omega_{11t} & \rho\left(\omega_{11t}\omega_{22t}\right)^{1/2} \\ \rho\left(\omega_{11t}\omega_{22t}\right)^{1/2} & \omega_{22t} \end{bmatrix} \tag{3.33}$$

where the variances ω_{iit} are time-varying for $i = 1, 2$. The correlation coefficient is constant over time, and we set $\rho = -0.9, 0.0$ and 0.9 to see how ρ affects the power of the test. The time-varying variances ω_{11t} and ω_{22t} are generated by the three parametric variance specifications discussed above. The parameters in (3.5), (3.6), (3.23), and (3.28) are defined as follows:

Parametric cases:	
White specification	
$n = 2, 4,$	$\delta_1 = \delta_2 = \frac{1}{j}(1, 1, \ldots, 1)', \quad j = 1, 10, 100$
ARCH specification	
$q = 1, 3, 5,$	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.25, 0.20, 0.15, 0.10, 0.05)$
Smooth transition specification with $\tau_{it} = t/T$	
$\gamma = 1, 10, 100,$	$c = \frac{T+1}{2T}, \ \lambda_1 = \lambda_2 = 1$

(3.34)

Furthermore, we set $\sigma_1^2 = \sigma_2^2 = 1$ in (3.2). In addition, we simulate the following three alternatives:

Other alternative cases:
1. Monotonically increasing variance
$\omega_{iit} = 1 + \frac{t-1}{T-1}(c-1)$
2. Step change in variance
$\omega_{iit} = 1 + (c-1)\, I\left(t > \left[\frac{T}{2}\right]\right)$
3. Random variance
$\omega_{iit} \sim U(1, c)$

(3.35)

In this setup, $c > 1$ and $I\left(t > \left[\frac{T}{2}\right]\right)$ is an indicator function: $I(A) = 1$ when A is true and $I(A) = 0$ when A is false. In the first case the variances change monotonically from 1 to c, the second case involves a structural change or a sudden increase of the variance, and in the third case the variance is completely stochastic. More specifically, c is set to $c = 2, 10$, where $c = 10$ implies a strong increase in the variances over time.

We thus have six possible variance specifications to test against. When generating the variances, we have three parametric families of time-varying variances with a total of twelve designs and three other cases with six possible designs in total. Including the three correlations, there are 324 possible alternatives. As a consequence only a small selection of the simulation results will be presented. The remaining ones are available from the authors upon request.

Note that the results are not corrected for any size distortion. This choice can be debated, but since tests such as this one are in practice usually applied without size-correction and because size distortion appears small, the size has not been adjusted. Tables 3.5-3.13 show the simulation results for a number of combinations. Consider first the results from the parametric cases (3.34) in Tables 3.5-3.10. The results in Tables 3.5-3.7 indicate, not surprisingly, that the test has high power against the correct family of time-varying variances. As would be expected, the power increases with the sample size but does it also when the time-variation becomes stronger.

We have also simulated tests against incorrect alternatives. The results in Table 3.8 show that testing against the ARCH(q) family when the variances are generated by the smooth transition specification, or vice versa, results in moderate to high power, between 0.50 and 0.95, if the time-variation is strong, as in the smooth transition case when $\gamma = 10$ or 100. Similar results are obtained for the other ARCH alternatives. Testing against the White family assuming that the alternative process is either an ARCH(q), Table 3.9, or having smoothly varying variances indicate that the tests have low or no power at all. From Table 3.10 it can be seen that the same result holds when the variances are generated by a member of the White family when testing against the ARCH(q) or a member of the smooth transition family.

For the three other alternative cases (3.35), the simulations yield results, in Tables 3.11-3.14, in line with the parametric ones. For the first two cases, with a monotonically increasing and a sudden increase of the variance in Tables 3.11 and 3.12, tests against the smooth transition alternative show high power, against the ARCH(q) case moderate to high power, and against the White alternative low power. The third case with a completely stochastic variance in Table 3.13 low and very slowly increasing power for all three alternatives considered in these simulations.

Finally, in none of the simulation alternatives did a correlation $|\rho| = 0.9$ lead to any major changes in power of the test compared to $\rho = 0$. Also, the joint test has higher power than the univariate tests in any of the combinations considered, given that both equations in the model have time-varying

variances. This is an expected result, since a joint test takes the dependence between the error processes into account. However, if only one of the equations in the model has a time-varying variance the joint test does not have as high power as the single-equation test. Carrying out both the joint test and the single-equation tests can therefore be recommended in practice.

3.7 Conclusions

In this paper we have derived a test of a constant error covariance matrix of a multivariate model under the assumption that the correlations remain constant even under the alternative. It is best viewed as a useful diagnostic test. A rejection of the null hypothesis against a particular parametric alternative does not necessarily mean that the true model is the specified one (3.1) with errors following this alternative. It could as well be an indication on a misspecified conditional mean or caused by a structural change.

The Monte Carlo simulations show that the LM test has good size properties. No major size distortion is detected in any of the experiments considered. The power simulations show that the test has very good power against a correctly specified alternative, but low or only up to moderate power in cases for a misspecified alternative hypothesis. Simulations has also been carried out in the case where the assumption of constant correlations is violated. The results show that the test is not in any major extent affected by time-varying correlations, except for a very particular case where the correlation increase monotonically over time and the variance is assumed to change smoothly with time under the alternative.

The assumption of constant correlations may be considered overly restrictive to begin with. Nevertheless, the results in this paper show that the assumption is a practical way of reducing the dimension of the problem of testing constancy of an error covariance matrix to a manageable size. The test may therefore be expected to be a useful addition to the toolbox of an applied econometrician.

Appendix A

Proof of Lemma 1

A.1 Derivation of average score vector and population information matrix.

The quasi-log likelihood of model (3.1) with covariance matrix (3.7) for time t is

$$\ell_t(\psi, \theta) = c + \ln\left|\mathbf{D}_t^{-1}\right| - \frac{1}{2}\ln|\mathbf{P}| - \frac{1}{2}\varepsilon_t'\mathbf{\Omega}_t^{-1}\varepsilon_t,$$

where θ is the vector of parameters in the covariance matrix $\mathbf{\Omega}_t$. Then

$$\frac{\partial \ell_t(\psi, \theta)}{\partial \theta} = \frac{\partial \ln\left|\mathbf{D}_t^{-1}\right|}{\partial \theta} - \frac{1}{2}\frac{\partial \ln|\mathbf{P}|}{\partial \theta} - \frac{1}{2}\frac{\partial \varepsilon_t'\mathbf{\Omega}_t^{-1}\varepsilon_t}{\partial \theta} \tag{A.1}$$

$$\begin{aligned}
= & \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \theta}\frac{\partial \ln\left|\mathbf{D}_t^{-1}\right|}{\partial vec\left(\mathbf{D}_t^{-1}\right)} \\
& -\frac{1}{2}\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \theta}\frac{\partial vec\left(\mathbf{D}_t^{-1}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\right)'}{\partial vec\left(\mathbf{D}_t^{-1}\right)}\frac{\partial\left(\varepsilon_t'\mathbf{\Omega}_t^{-1}\varepsilon_t\right)}{\partial vec\left(\mathbf{\Omega}_t^{-1}\right)} - \frac{1}{2}\frac{\partial vec(\mathbf{P})'}{\partial \theta}\frac{\partial \ln|\mathbf{P}|}{\partial vec(\mathbf{P})} \\
& -\frac{1}{2}\frac{\partial vec(\mathbf{P})'}{\partial \theta}\frac{\partial vec\left(\mathbf{P}^{-1}\right)'}{\partial vec(\mathbf{P})}\frac{\partial vec\left(\mathbf{D}_t^{-1}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\right)'}{\partial vec\left(\mathbf{P}^{-1}\right)}\frac{\partial\left(\varepsilon_t'\mathbf{\Omega}_t^{-1}\varepsilon_t\right)}{\partial vec\left(\mathbf{\Omega}_t^{-1}\right)} \\
= & \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \theta}\left(vec(\mathbf{D}_t) - \frac{1}{2}\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\otimes\mathbf{I}_m + \mathbf{I}_m\otimes\mathbf{P}^{-1}\mathbf{D}_t^{-1}\right)vec\left(\varepsilon_t\varepsilon_t'\right)\right) \\
& +\frac{1}{2}\frac{\partial vec(\mathbf{P})'}{\partial \theta}\left(\left(\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\right)\left(\mathbf{D}_t^{-1}\otimes\mathbf{D}_t^{-1}\right)vec\left(\varepsilon_t\varepsilon_t'\right) - vec\left(\mathbf{P}^{-1}\right)\right)
\end{aligned}$$

$$= \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{D}_t - \frac{1}{2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \frac{1}{2}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)$$
$$+\frac{1}{2}\frac{\partial vec\left(\mathbf{P}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \mathbf{P}^{-1}\right)$$

where

$$\frac{\partial \ln\left|\mathbf{D}_t^{-1}\right|}{\partial vec\left(\mathbf{D}_t^{-1}\right)} = vec\left(\mathbf{D}_t\right)$$
$$\frac{\partial vec\left(\mathbf{D}_t^{-1}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\right)'}{\partial vec\left(\mathbf{D}_t^{-1}\right)} = \mathbf{P}^{-1}\mathbf{D}_t^{-1} \otimes \mathbf{I}_m + \mathbf{I}_m \otimes \mathbf{P}^{-1}\mathbf{D}_t^{-1}$$
$$\frac{\partial\left(\boldsymbol{\varepsilon}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\varepsilon}_t\right)}{\partial vec\left(\boldsymbol{\Omega}_t^{-1}\right)} = vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)$$
$$\frac{\partial \ln|\mathbf{P}|}{\partial vec\left(\mathbf{P}\right)} = vec\left(\mathbf{P}^{-1}\right)$$
$$\frac{\partial vec\left(\mathbf{P}^{-1}\right)'}{\partial vec\left(\mathbf{P}\right)} = -\left(\mathbf{P}^{-1} \otimes \mathbf{P}^{-1}\right)$$
$$\frac{\partial vec\left(\mathbf{D}_t^{-1}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\right)'}{\partial vec\left(\mathbf{P}^{-1}\right)} = \mathbf{D}_t^{-1} \otimes \mathbf{D}_t^{-1}$$

The average score vector is then given by

$$\bar{\mathbf{s}}_T\left(\boldsymbol{\theta}\right) = \frac{1}{T}\sum_1^T \left\{ \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{D}_t - \frac{1}{2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \frac{1}{2}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\right.$$
$$\left. +\frac{1}{2}\frac{\partial vec\left(\mathbf{P}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \mathbf{P}^{-1}\right)\right\}. \tag{A.2}$$

The population information matrix is defined by the second order derivatives

$$\frac{\partial^2 \ell_t\left(\boldsymbol{\psi}, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}'}\left(\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{D}_t - \frac{1}{2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \frac{1}{2}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\right.$$
$$\left. +\frac{1}{2}\frac{\partial vec\left(\mathbf{P}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} - \mathbf{P}^{-1}\right)\right).$$

Taking the terms one by one

Term 1

$$\frac{\partial}{\partial \boldsymbol{\theta}'}\left(\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{D}_t\right)\right) = \tag{A.3}$$

$$
\begin{aligned}
&= \left(vec\left(\mathbf{D}_t\right)' \otimes \mathbf{I}_p\right) \frac{\partial}{\partial \boldsymbol{\theta}'} vec\left(\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}}\right) + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \frac{\partial vec\left(\mathbf{D}_t\right)}{\partial \boldsymbol{\theta}'} \\
&= \left(vec\left(\mathbf{D}_t\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \frac{\partial vec\left(\mathbf{D}_t\right)}{\partial vec\left(\mathbf{D}_t^{-1}\right)'} \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \\
&= \left(vec\left(\mathbf{D}_t\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} - \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\mathbf{D}_t \otimes \mathbf{D}_t\right) \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'}
\end{aligned}
$$

Term 2

$$
\frac{\partial}{\partial \boldsymbol{\theta}'} \left(\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} vec\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{D}_t^{-1} \mathbf{P}^{-1}\right)\right) = \tag{A.4}
$$

$$
\begin{aligned}
&= \left(vec\left(\mathbf{P}^{-1} \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right)' \otimes \mathbf{I}_p\right) \frac{\partial}{\partial \boldsymbol{\theta}'} vec\left(\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}}\right) \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \frac{\partial vec\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{D}_t^{-1} \mathbf{P}^{-1}\right)}{\partial \boldsymbol{\theta}'} \\
&= \left(vec\left(\mathbf{P}^{-1} \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\frac{\partial vec\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{D}_t^{-1} \mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{D}_t^{-1}\right)'} \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \right. \\
&\quad \left. + \frac{\partial vec\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{D}_t^{-1} \mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{P}^{-1}\right)'} \frac{\partial vec\left(\mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{P}\right)'} \frac{\partial vec\left(\mathbf{P}\right)}{\partial \boldsymbol{\theta}'} \right) \\
&= \left(vec\left(\mathbf{P}^{-1} \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\left(\mathbf{P}^{-1} \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right) \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \right. \\
&\quad \left. - \left(\mathbf{I}_m \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{D}_t^{-1}\right) \left(\mathbf{P}^{-1} \otimes \mathbf{P}^{-1}\right) \frac{\partial vec\left(\mathbf{P}\right)}{\partial \boldsymbol{\theta}'} \right) \\
&= \left(vec\left(\mathbf{P}^{-1} \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\mathbf{P}^{-1} \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right) \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \\
&\quad - \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\mathbf{P}^{-1} \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{D}_t^{-1} \mathbf{P}^{-1}\right) \frac{\partial vec\left(\mathbf{P}\right)}{\partial \boldsymbol{\theta}'}
\end{aligned}
$$

Term 3

$$\frac{\partial}{\partial \boldsymbol{\theta}'} = \left(\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) \right) \tag{A.5}$$

$$\begin{aligned}
&= \left(vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)' \otimes \mathbf{I}_p\right) \frac{\partial}{\partial \boldsymbol{\theta}'} vec\left(\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \right) \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \frac{\partial vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)}{\partial \boldsymbol{\theta}'} \\
&= \left(vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}'\partial \boldsymbol{\theta}} \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\frac{\partial vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)}{\partial vec\left(\mathbf{D}_t^{-1}\right)'} \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \right. \\
&\quad \left. + \frac{\partial vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)}{\partial vec\left(\mathbf{P}^{-1}\right)'} \frac{\partial vec\left(\mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{P}\right)'} \frac{\partial vec\left(\mathbf{P}\right)}{\partial \boldsymbol{\theta}'} \right) \\
&= \left(vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}'\partial \boldsymbol{\theta}} \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \otimes \mathbf{P}^{-1}\right) \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \right. \\
&\quad \left. - \left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1} \otimes \mathbf{I}_m\right)\left(\mathbf{P}^{-1} \otimes \mathbf{P}^{-1}\right) \frac{\partial vec\left(\mathbf{P}\right)}{\partial \boldsymbol{\theta}'} \right) \\
&= \left(vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)' \otimes \mathbf{I}_p\right) \frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}'\partial \boldsymbol{\theta}} \\
&\quad + \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \otimes \mathbf{P}^{-1}\right) \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \\
&\quad - \frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial \boldsymbol{\theta}} \left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1} \otimes \mathbf{P}^{-1}\right) \frac{\partial vec\left(\mathbf{P}\right)}{\partial \boldsymbol{\theta}'}
\end{aligned}$$

Term 4

$$\frac{\partial}{\partial \boldsymbol{\theta}'} \left(\frac{\partial vec\left(\mathbf{P}\right)'}{\partial \boldsymbol{\theta}} vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right) \right) = \tag{A.6}$$

$$\begin{aligned}
&= \left(vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)' \otimes \mathbf{I}_p\right) \frac{\partial}{\partial \boldsymbol{\theta}'} vec\left(\frac{\partial vec\left(\mathbf{P}\right)'}{\partial \boldsymbol{\theta}} \right) \\
&\quad + \frac{\partial vec\left(\mathbf{P}\right)'}{\partial \boldsymbol{\theta}} \frac{\partial vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)}{\partial \boldsymbol{\theta}'}
\end{aligned}$$

$$
\begin{aligned}
&= \left(vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)'\otimes\mathbf{I}_p\right)\frac{\partial^2 vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}} \\
&\quad+\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\left(\frac{\partial vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{D}_t^{-1}\right)'}\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial\boldsymbol{\theta}'}\right) \\
&\quad+\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\left(\frac{\partial vec\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{P}^{-1}\right)'}\frac{\partial vec\left(\mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{P}\right)'}\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'}\right) \\
&= \frac{\partial vec(\mathbf{P})'}{\partial\boldsymbol{\theta}}\Big[\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\otimes\mathbf{P}^{-1}+\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial\boldsymbol{\theta}'} \\
&\quad-\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\otimes\mathbf{I}_m+\mathbf{I}_m\otimes\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\right)\times \\
&\quad\times\left(\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\right)\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'}\Big] \\
&= \frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\otimes\mathbf{P}^{-1}+\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial\boldsymbol{\theta}'} \\
&\quad-\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\right. \\
&\quad\left.+\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'},
\end{aligned}
$$

$$\text{as}\quad\frac{\partial^2 vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}=0.$$

Term 5

$$\frac{\partial}{\partial\boldsymbol{\theta}'}\left(\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}vec\left(\mathbf{P}^{-1}\right)\right)= \tag{A.7}$$

$$
\begin{aligned}
&= \left(vec\left(\mathbf{P}^{-1}\right)'\otimes\mathbf{I}_p\right)\frac{\partial}{\partial\boldsymbol{\theta}'}vec\left(\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\right)+\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\frac{\partial vec\left(\mathbf{P}^{-1}\right)}{\partial\boldsymbol{\theta}'} \\
&= \left(vec\left(\mathbf{P}^{-1}\right)'\otimes\mathbf{I}_p\right)\frac{\partial^2 vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}+\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\frac{\partial vec\left(\mathbf{P}^{-1}\right)}{\partial vec\left(\mathbf{P}\right)'}\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'} \\
&= -\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\left(\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\right)\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'},
\end{aligned}
$$

$$\text{as}\quad\frac{\partial^2 vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}=0.$$

Then taking the sum over all t, the average Hessian matrix becomes

$$
\begin{aligned}
\overline{\mathbf{H}}_T(\boldsymbol{\theta}) &= -\frac{1}{T}\sum_{1}^{T}\frac{\partial^2 l_t}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}} \qquad (A.8)\\
&= \frac{1}{T}\sum_{1}^{T}\left\{\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial\boldsymbol{\theta}}\left\{\mathbf{D}_t\otimes\mathbf{D}_t+\right.\right.\\
&\quad \left.\left.+\frac{1}{2}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\otimes\mathbf{P}^{-1}\right)+\frac{1}{2}\left(\mathbf{P}^{-1}\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\right\}\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial\boldsymbol{\theta}'}\right.\\
&\quad -\frac{\partial^2 vec\left(\mathbf{D}_t^{-1}\right)'}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}\left(\mathbf{I}_p\otimes vec\left(\mathbf{D}_t-\frac{1}{2}\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'-\frac{1}{2}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)\right)\\
&\quad -\frac{1}{2}\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial\boldsymbol{\theta}}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}+\mathbf{P}^{-1}\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\right)\frac{\partial vec(\mathbf{P})}{\partial\boldsymbol{\theta}'}\\
&\quad -\frac{1}{2}\frac{\partial vec(\mathbf{P})}{\partial\boldsymbol{\theta}'}\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\otimes\mathbf{P}^{-1}+\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial\boldsymbol{\theta}'}\\
&\quad +\frac{1}{2}\frac{\partial vec(\mathbf{P})'}{\partial\boldsymbol{\theta}}\left(\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}+\right.\\
&\quad \left.\left.+\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-1}\mathbf{P}^{-1}-\left(\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\right)\right)\frac{\partial vec(\mathbf{P})}{\partial\boldsymbol{\theta}'}\right\}
\end{aligned}
$$

Then using the law of iterated expectations, given the information set I_t, $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\, I_t\right)=\boldsymbol{\Omega}_t=\mathbf{D}_t\mathbf{P}\mathbf{D}_t$, the population information matrix is given by

$$
\begin{aligned}
\mathbf{I}_T(\boldsymbol{\theta}) &= \frac{1}{T}\sum_{1}^{T}E\left\{\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial\boldsymbol{\theta}}\left\{\mathbf{D}_t\otimes\mathbf{D}_t+\right.\right.\\
&\quad \left.\left.+\frac{1}{2}\left(\boldsymbol{\Omega}_t\otimes\mathbf{P}^{-1}\right)+\frac{1}{2}\left(\mathbf{P}^{-1}\otimes\boldsymbol{\Omega}_t\right)\right\}\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial\boldsymbol{\theta}'}\right.\\
&\quad -\frac{1}{2}\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)'}{\partial\boldsymbol{\theta}}\left(\mathbf{D}_t\otimes\mathbf{P}^{-1}+\mathbf{P}^{-1}\otimes\mathbf{D}_t\right)\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'}\\
&\quad -\frac{1}{2}\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'}\left(\mathbf{P}^{-1}\otimes\mathbf{D}_t+\mathbf{D}_t\otimes\mathbf{P}^{-1}\right)\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial\boldsymbol{\theta}'}\\
&\quad \left.+\frac{1}{2}\frac{\partial vec\left(\mathbf{P}\right)'}{\partial\boldsymbol{\theta}}\left(\mathbf{P}^{-1}\otimes\mathbf{P}^{-1}\right)\frac{\partial vec\left(\mathbf{P}\right)}{\partial\boldsymbol{\theta}'}\right\}
\end{aligned}
$$

where p in $\mathbf{I}_p$ is the number of parameters in $\boldsymbol{\theta}$. This concludes the proof. ■

A.2 Expression of first order partial derivatives of $\mathbf{D}_t^{-1}$.

Note that the matrix $\mathbf{D}_t^{-1}$ can be written as

$$\mathbf{D}_t^{-1} = \operatorname{diag}\left(H\left(\boldsymbol{\varphi}_1\right), H\left(\boldsymbol{\varphi}_2\right), \ldots, H\left(\boldsymbol{\varphi}_m\right)\right),$$

where $\boldsymbol{\varphi}_i$ are the parameters included in the variance ω_{iit}, $\boldsymbol{\theta} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \ldots, \boldsymbol{\varphi}_m', \boldsymbol{\rho}')'$, and

$$H\left(\boldsymbol{\varphi}_i\right) = \frac{1}{\left[\omega_{iit}\left(\boldsymbol{\varphi}_i\right)\right]^{1/2}} = \frac{1}{\left[h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right)\right]^{1/2}}, \qquad i = 1, \ldots, m. \tag{A.9}$$

As before, let p denote the total number of parameters in $\boldsymbol{\theta}$, $p_i + 1$ the number in $\boldsymbol{\varphi}_i$, not necessarily equal for all i, and p_ρ the number in $\boldsymbol{\rho}$. After reparameterizing, if necessary, the first order partial derivatives are given by

$$\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} = \tag{A.10}$$

$$= \begin{bmatrix} \frac{\partial H(\varphi_1)}{\partial h_1(\varphi_1'\mathbf{v}_{1t})}\left(\frac{\partial h_1(\varphi_1'\mathbf{v}_{1t})}{\partial \varphi_1'}, \ \cdots, \ \frac{\partial h_1(\varphi_1'\mathbf{v}_{1t})}{\partial \varphi_i'}, \ \cdots, \ \frac{\partial h_1(\varphi_1'\mathbf{v}_{1t})}{\partial \varphi_m'}, \ \frac{\partial h_1(\varphi_1'\mathbf{v}_{1t})}{\partial \rho'}\right) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \frac{\partial H(\varphi_i)}{\partial h_i(\varphi_i'\mathbf{v}_{it})}\left(\frac{\partial h_i(\varphi_i'\mathbf{v}_{it})}{\partial \varphi_1'}, \ \cdots, \ \frac{\partial h_i(\varphi_i'\mathbf{v}_{it})}{\partial \varphi_i'}, \ \cdots, \ \frac{\partial h_i(\varphi_i'\mathbf{v}_{it})}{\partial \varphi_m'}, \ \frac{\partial h_i(\varphi_i'\mathbf{v}_{it})}{\partial \rho'}\right) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \frac{\partial H(\varphi_m)}{\partial h_m(\varphi_m'\mathbf{v}_{mt})}\left(\frac{\partial h_m(\varphi_m'\mathbf{v}_{mt})}{\partial \varphi_1'}, \ \cdots, \ \frac{\partial h_m(\varphi_m'\mathbf{v}_{mt})}{\partial \varphi_i'}, \ \cdots, \ \frac{\partial h_m(\varphi_m'\mathbf{v}_{mt})}{\partial \varphi_m'}, \ \frac{\partial h_m(\varphi_m'\mathbf{v}_{mt})}{\partial \rho'}\right) \end{bmatrix}$$

$$= \operatorname{diag}\left(-\frac{1}{2\left[h_1\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)\right]^{3/2}}\frac{\partial h_1\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)}{\partial \boldsymbol{\varphi}_1'\mathbf{v}_{1t}}, 0, \ldots, 0, -\frac{1}{2\left[h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right)\right]^{3/2}}\frac{\partial h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right)}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}},\right.$$

$$\left. 0, \ldots, 0, -\frac{1}{2\left[h_m\left(\boldsymbol{\varphi}_m'\mathbf{v}_{mt}\right)\right]^{3/2}}\frac{\partial h_m\left(\boldsymbol{\varphi}_m'\mathbf{v}_{mt}\right)}{\partial \boldsymbol{\varphi}_m'\mathbf{v}_{mt}}\right) \times$$

$$\times \begin{bmatrix} \frac{\partial \boldsymbol{\varphi}_1' \mathbf{v}_{1t}}{\partial \boldsymbol{\varphi}_1'} & \dots & \mathbf{0}_i & \dots & \mathbf{0}_m & \mathbf{0}_\rho \\ \mathbf{0}_1 & \dots & \mathbf{0}_i & \dots & \mathbf{0}_m & \mathbf{0}_\rho \\ \vdots & \dots & \vdots & \dots & \vdots & \vdots \\ \mathbf{0}_1 & \dots & \mathbf{0}_i & \dots & \mathbf{0}_m & \mathbf{0}_\rho \\ \mathbf{0}_1 & \dots & \frac{\partial \boldsymbol{\varphi}_i' \mathbf{v}_{it}}{\partial \boldsymbol{\varphi}_i'} & \dots & \mathbf{0}_m & \mathbf{0}_\rho \\ \mathbf{0}_1 & \dots & \mathbf{0}_i & \dots & \mathbf{0}_m & \mathbf{0}_\rho \\ \vdots & \dots & \vdots & \dots & \vdots & \vdots \\ \mathbf{0}_1 & \dots & \mathbf{0}_i & \dots & \mathbf{0}_m & \mathbf{0}_\rho \\ \mathbf{0}_1 & \dots & \mathbf{0}_i & \dots & \frac{\partial \boldsymbol{\varphi}_m' \mathbf{v}_{mt}}{\partial \boldsymbol{\varphi}_m'} & \mathbf{0}_\rho \end{bmatrix}$$

since $\frac{\partial H(\boldsymbol{\varphi}_i)}{\partial h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it})} = -\frac{1}{2h_i^{3/2}(\boldsymbol{\varphi}_i' \mathbf{v}_{it})}$, $\frac{\partial h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it})}{\partial \boldsymbol{\rho}'} = \mathbf{0}$ and $\frac{\partial h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it})}{\partial \boldsymbol{\varphi}_j'} = \mathbf{0}_j$ for $i, j = 1, \dots, n$, $i \neq j$. Note that $\mathbf{0}_i$ is a $(1 \times (p_i + 1))$ null vector, $\mathbf{0}_\rho$ is a $(1 \times p_\rho)$ null vector, the diagonal matrix is $(m^2 \times m^2)$ and the second matrix is $(m^2 \times p)$. Note also that $\frac{\partial h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it})}{\partial \boldsymbol{\varphi}_i'}$ is a $(1 \times (p_i + 1))$ vector and $\frac{\partial h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it})}{\partial \boldsymbol{\rho}'} = \mathbf{0}_\rho$ is a $(1 \times p_\rho)$ vector.

A.2.1 Special case 1. $h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it}) = h_i(\sigma_i^2 + \boldsymbol{\delta}_i' vech(\mathbf{x}_t \mathbf{x}_t'))$, $i = 1, 2$.

Inserting $h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it})$ into (A.9), the function $H(\boldsymbol{\varphi}_i)$ equals,

$$H(\boldsymbol{\varphi}_i) = \frac{1}{[h_i(\boldsymbol{\varphi}_i' \mathbf{v}_{it})]^{1/2}} = \frac{1}{\left[h_i\left(\sigma_i^2 + \boldsymbol{\delta}_i' vech(\mathbf{x}_t \mathbf{x}_t')\right)\right]^{1/2}}, \; i = 1, 2.$$

The first order partial derivatives then follows from (A.10) as

$$\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} = \operatorname{diag}\left(-\frac{h_1'(\boldsymbol{\varphi}_1' \mathbf{v}_{1t})}{2[h_1(\boldsymbol{\varphi}_1' \mathbf{v}_{1t})]^{3/2}}, 0, 0, -\frac{h_2'(\boldsymbol{\varphi}_2' \mathbf{v}_{2t})}{2[h_2(\boldsymbol{\varphi}_2' \mathbf{v}_{2t})]^{3/2}}\right) \times \begin{bmatrix} \frac{\partial \boldsymbol{\varphi}_1' \mathbf{v}_{1t}}{\partial \boldsymbol{\varphi}_1'} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{\partial \boldsymbol{\varphi}_2' \mathbf{v}_{2t}}{\partial \boldsymbol{\varphi}_2'} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{h_1'(\varphi_1'\mathbf{v}_{1t})}{2\left[h_1\left(\sigma_1^2+\boldsymbol{\delta}_1' vech(\mathbf{x}_t\mathbf{x}_t')\right)\right]^{3/2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{h_2'(\varphi_2'\mathbf{v}_{2t})}{2\left[h_2\left(\sigma_2^2+\boldsymbol{\delta}_2' vech(\mathbf{x}_t\mathbf{x}_t')\right)\right]^{3/2}} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & vech\left(\mathbf{x}_t\mathbf{x}_t'\right)' & 0 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 1 & vech\left(\mathbf{x}_t\mathbf{x}_t'\right)' & 0 \end{bmatrix},$$

where $\mathbf{0}$ is a $(1 \times n(n+1)/2)$ null vector, that is the same number of elements as in $vech\left(\mathbf{x}_t\mathbf{x}_t'\right)'$, and $h_i'\left(\varphi_i'\mathbf{v}_{it}\right) = \frac{\partial h_i\left(\varphi_i'\mathbf{v}_{it}\right)}{\partial \varphi_i'\mathbf{v}_{it}}$, $i = 1, 2$. Estimated under the null,

$$\left.\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'}\right|_{H_0} = \begin{bmatrix} -\frac{h_1'(\varphi_1'\mathbf{v}_{1t})|_{H_0}}{2\left[h_1\left(\tilde{\sigma}_1^2\right)\right]^{3/2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{h_2'(\varphi_2'\mathbf{v}_{2t})|_{H_0}}{2\left[h_2\left(\tilde{\sigma}_2^2\right)\right]^{3/2}} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & vech\left(\mathbf{x}_t\mathbf{x}_t'\right)' & 0 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 0 & \mathbf{0} & 0 \\ 0 & \mathbf{0} & 1 & vech\left(\mathbf{x}_t\mathbf{x}_t'\right)' & 0 \end{bmatrix}$$

A.2.2 Special case 2. $h_i\left(\varphi_i'\mathbf{v}_{it}\right) = h_i\left(\sigma_i^2 + \sum_{j=1}^{q} \alpha_{ij}\varepsilon_{it-j}^2\right)$, $i = 1, 2$.

After inserting the variance specification in (A.9) we have

$$H\left(\varphi_i\right) = \frac{1}{\left[h_i\left(\varphi_i'\mathbf{v}_{it}\right)\right]^{1/2}} = \frac{1}{\left[h_i\left(\sigma_i^2 + \alpha_{i1}\varepsilon_{it-1}^2 + \ldots + \alpha_{in}\varepsilon_{it-q}^2\right)\right]^{1/2}}, \; i = 1, 2.$$

The first order partial derivatives then follows

$$\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} = \operatorname{diag}\left(-\frac{h_1'\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)}{2\left[h_1\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)\right]^{3/2}}, 0, 0, -\frac{h_2'\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)}{2\left[h_2\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)\right]^{3/2}}\right) \times \begin{bmatrix} \frac{\partial \boldsymbol{\varphi}_1'\mathbf{v}_{1t}}{\partial \boldsymbol{\varphi}_1'} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{\partial \boldsymbol{\varphi}_2'\mathbf{v}_{2t}}{\partial \boldsymbol{\varphi}_2'} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{h_1'\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)}{2\left[h_1\left(\sigma_1^2+\alpha_{11}\varepsilon_{1t-1}^2+\ldots+\alpha_{1n}\varepsilon_{1t-q}^2\right)\right]^{3/2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{h_2'\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)}{2\left[h_2\left(\sigma_2^2+\alpha_{21}\varepsilon_{2t-1}^2+\ldots+\alpha_{2n}\varepsilon_{2t-q}^2\right)\right]^{3/2}} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & \varepsilon_{1t-1}^2 & \ldots & \varepsilon_{1t-q}^2 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 & \varepsilon_{2t-1}^2 & \ldots & \varepsilon_{2t-q}^2 & 0 \end{bmatrix},$$

where $h_i'\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right) = \frac{\partial h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right)}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}$, $i = 1, 2$ as before. Estimated under the null the expression equals

$$\left.\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'}\right|_{H_0} = \begin{bmatrix} -\frac{h_1'\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)\big|_{H_0}}{2\left[h_1\left(\tilde{\sigma}_1^2\right)\right]^{3/2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{h_2'\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)\big|_{H_0}}{2\left[h_2\left(\tilde{\sigma}_2^2\right)\right]^{3/2}} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & \tilde{\varepsilon}_{1t-1}^2 & \ldots & \tilde{\varepsilon}_{1t-q}^2 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 & \tilde{\varepsilon}_{2t-1}^2 & \ldots & \tilde{\varepsilon}_{2t-q}^2 & 0 \end{bmatrix}.$$

A.2.3 Special case 3. $h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right) = h_i\left(t/T\right)$, $i = 1, 2$.

Inserting $h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right) = h_i\left(t/T\right)$ into (A.9), no reparameterizing needed for this case, then $H\left(\boldsymbol{\varphi}_i\right) = \frac{1}{\left[h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right)\right]^{1/2}} = \frac{1}{\left[h_i\left(\varsigma_i^2+\lambda_i^* t/T+\alpha_i R_i\right)\right]^{1/2}}$, $i = 1, 2$, and (A.10)

follows as

$$\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'} = \operatorname{diag}\left(-\frac{h_1'\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)}{2\left[h_1\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)\right]^{3/2}}, 0, , 0, -\frac{h_2'\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)}{2\left[h_2\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)\right]^{3/2}}\right) \times \begin{bmatrix} \frac{\partial \boldsymbol{\varphi}_1'\mathbf{v}_{1t}}{\partial \boldsymbol{\varphi}_1'} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{\partial \boldsymbol{\varphi}_2'\mathbf{v}_{2t}}{\partial \boldsymbol{\varphi}_2'} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{h_1'\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)}{2\left[h_1\left(\varsigma_1^2+\lambda_1^* t/T+\alpha_1 R_1\right)\right]^{3/2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{h_2'\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)}{2\left[h_2\left(\varsigma_2^2+\lambda_2^* t/T+\alpha_2 R_2\right)\right]^{3/2}} \end{bmatrix} \times \begin{bmatrix} 1 & t/T & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & t/T & 0 \end{bmatrix},$$

where $h_i'\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right) = \frac{\partial h_i\left(\boldsymbol{\varphi}_i'\mathbf{v}_{it}\right)}{\partial \boldsymbol{\varphi}_i'\mathbf{v}_{it}}$, $i = 1, 2$ as above. The derivative becomes after estimated under the null hypothesis

$$\left.\frac{\partial vec\left(\mathbf{D}_t^{-1}\right)}{\partial \boldsymbol{\theta}'}\right|_{H_0} = \begin{bmatrix} -\frac{h_1'\left(\boldsymbol{\varphi}_1'\mathbf{v}_{1t}\right)\big|_{H_0}}{2\left[h_1\left(\tilde{\sigma}_1^2\right)\right]^{3/2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{h_2'\left(\boldsymbol{\varphi}_2'\mathbf{v}_{2t}\right)\big|_{H_0}}{2\left[h_2\left(\tilde{\sigma}_2^2\right)\right]^{3/2}} \end{bmatrix} \times \begin{bmatrix} 1 & t/T & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & t/T & 0 \end{bmatrix}.$$

Appendix B

Tables

B.1 Size simulations

Table 3.1 Empirical size of the test of constant variance (3.13) against the 'White specification' with four explanatory variables, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test.

	$\rho = -0.9$			$\rho = 0.0$			$\rho = 0.9$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.067	0.067	0.070	0.068	0.067	0.069	0.069	0.067	0.070
50	0.075	0.076	0.082	0.078	0.076	0.083	0.077	0.077	0.082
100	0.076	0.074	0.082	0.075	0.075	0.081	0.075	0.076	0.082
250	0.067	0.066	0.073	0.067	0.068	0.072	0.069	0.068	0.073
500	0.062	0.062	0.064	0.061	0.063	0.064	0.061	0.061	0.064
1000	0.057	0.056	0.059	0.058	0.057	0.060	0.058	0.058	0.061

Table 3.2 Empirical size of the test of constant variance (3.13) against the ARCH(1) specification, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test.

	$\rho = -0.9$			$\rho = 0.0$			$\rho = 0.9$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.013	0.012	0.014	0.012	0.013	0.012	0.012	0.012	0.015
50	0.022	0.022	0.025	0.022	0.022	0.023	0.022	0.023	0.025
100	0.032	0.033	0.034	0.031	0.032	0.032	0.032	0.031	0.034
250	0.041	0.041	0.042	0.040	0.041	0.041	0.040	0.042	0.042
500	0.044	0.045	0.045	0.045	0.044	0.044	0.046	0.045	0.046
1000	0.047	0.048	0.048	0.047	0.047	0.048	0.048	0.047	0.048

Table 3.3 Empirical size of the test of constant variance (3.13) against the smooth transition specification with $\tau_{it} = t/T$, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test.

	$\rho = -0.9$			$\rho = 0.0$			$\rho = 0.9$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.041	0.041	0.038	0.041	0.040	0.036	0.041	0.040	0.037
50	0.044	0.045	0.043	0.045	0.045	0.043	0.046	0.046	0.043
100	0.049	0.048	0.047	0.046	0.047	0.046	0.049	0.048	0.046
250	0.049	0.048	0.049	0.050	0.050	0.049	0.050	0.049	0.050
500	0.051	0.050	0.050	0.049	0.050	0.048	0.049	0.051	0.050
1000	0.050	0.050	0.050	0.050	0.050	0.050	0.049	0.049	0.049

B.2 Simulating time-varying correlation

Table 3.4 Estimated rejection frequencies of the test (3.13) assuming constant variances but time-varying correlation ρ_t, $t = 1, \ldots, T$, at sample sizes 25, 50, 100, 250, 500 and 1000, when testing against:

White, $n = 4$		25	50	100	250	500	1000
1.	$\rho_t = -1, \ldots, 1$	0.060	0.080	0.084	0.078	0.072	0.067
2.	$\rho_t \sim U(-1, 1)$	0.062	0.082	0.084	0.079	0.073	0.067
ARCH(1)		25	50	100	250	500	1000
1.	$\rho_t = -1, \ldots, 1$	0.007	0.018	0.029	0.040	0.045	0.049
2.	$\rho_t \sim U(-1, 1)$	0.012	0.022	0.031	0.041	0.045	0.047
Smooth transition		25	50	100	250	500	1000
1.	$\rho_t = -1, \ldots, 1$	0.004	0.005	0.006	0.006	0.006	0.006
2.	$\rho_t \sim U(-1, 1)$	0.035	0.044	0.050	0.052	0.052	0.053

B.3 Power simulations

Table 3.5 Estimated rejection frequencies of the test of constant variance (3.13) against the 'White specification', 2 explanatory variables, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test. Error variances are generated by the 'White specification' with $n = 4$, $\rho = 0.9$, and parameters δ in (3.5) as:

	$\delta = (1, 1, \ldots, 1)/100$			$\delta = (1, 1, \ldots, 1)/10$			$\delta = (1, 1, \ldots, 1)$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.064	0.063	0.074	0.163	0.163	0.209	0.365	0.366	0.495
50	0.073	0.075	0.088	0.245	0.245	0.341	0.570	0.571	0.757
100	0.082	0.081	0.099	0.373	0.372	0.529	0.807	0.806	0.947
250	0.092	0.092	0.113	0.653	0.656	0.850	0.988	0.988	1.000
500	0.109	0.109	0.136	0.896	0.896	0.987	1.000	1.000	1.000
1000	0.140	0.140	0.184	0.994	0.994	1.000	1.000	1.000	1.000

Table 3.6 Estimated rejection frequencies of the test of constant variance (3.13) against the ARCH(3) specification, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test.
Error variances, for $\rho = 0.0$, are generated with parameters in (3.6) given by $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.25, 0.20, 0.15, 0.10, 0.05)$ for:

	ARCH(1)			ARCH(3)			ARCH(5)		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.056	0.056	0.064	0.089	0.089	0.109	0.088	0.085	0.108
50	0.159	0.160	0.222	0.314	0.314	0.454	0.327	0.327	0.472
100	0.340	0.341	0.498	0.635	0.635	0.837	0.673	0.674	0.869
250	0.704	0.703	0.896	0.958	0.959	0.998	0.971	0.971	0.999
500	0.940	0.939	0.996	0.999	0.999	1.000	1.000	1.000	1.000
1000	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.7 Estimated rejection frequencies of the test of constant variance (3.13) against the smooth transition specification, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test. Error variances are generated by the smooth transition specification, and $\rho = 0.9$.

	$\gamma = 1$			$\gamma = 10$			$\gamma = 100$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.045	0.045	0.042	0.195	0.195	0.241	0.248	0.249	0.326
50	0.059	0.057	0.058	0.436	0.434	0.602	0.555	0.554	0.751
100	0.074	0.074	0.081	0.767	0.768	0.934	0.880	0.883	0.983
250	0.122	0.123	0.152	0.993	0.993	1.000	0.999	0.999	1.000
500	0.202	0.202	0.270	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.354	0.356	0.505	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.8 Estimated rejection frequencies of the test of constant variance (3.13) against the ARCH(1) specification, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test. Error variances are generated by the smooth transition specification with $\tau_{it} = t/T$, $\rho = 0.0$.

	$\gamma = 1$			$\gamma = 10$			$\gamma = 100$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.014	0.013	0.013	0.039	0.041	0.045	0.055	0.055	0.063
50	0.024	0.024	0.025	0.073	0.075	0.092	0.110	0.109	0.141
100	0.035	0.035	0.036	0.117	0.116	0.155	0.184	0.184	0.256
250	0.043	0.045	0.045	0.205	0.206	0.289	0.348	0.345	0.496
500	0.048	0.049	0.051	0.326	0.327	0.469	0.551	0.557	0.753
1000	0.052	0.050	0.054	0.530	0.528	0.727	0.813	0.813	0.957

Table 3.9 Estimated rejection frequencies of the test of constant variance (3.13) against the 'White specification', 2 explanatory variables, at sample sizes 25, 50, 100, 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test. Error variances, for $\rho = 0.0$, are generated with parameters in (3.6) given by $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.25, 0.20, 0.15, 0.10, 0.05)$ for;

	$ARCH\,(1)$			$ARCH\,(3)$			$ARCH\,(5)$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.059	0.059	0.067	0.062	0.063	0.074	0.062	0.061	0.072
50	0.068	0.069	0.083	0.086	0.085	0.111	0.090	0.090	0.116
100	0.076	0.076	0.094	0.113	0.114	0.155	0.131	0.131	0.180
250	0.083	0.084	0.103	0.161	0.159	0.223	0.205	0.206	0.295
500	0.090	0.088	0.109	0.195	0.197	0.281	0.270	0.269	0.394
1000	0.091	0.093	0.115	0.229	0.232	0.335	0.337	0.338	0.492

Table 3.10 Estimated rejection frequencies of the test of constant variance (3.13) against the smooth transition specification, at sample sizes 25, 50, 100 250, 500 and 1000. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test.
Error variances generated by the 'White specification' with $n = 2$, $\rho = 0.0$.

	$\delta = (1, 1, \ldots, 1)/100$			$\delta = (1, 1, \ldots, 1)/10$			$\delta = (1, 1, \ldots, 1)$		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.040	0.040	0.037	0.048	0.048	0.045	0.100	0.098	0.110
50	0.044	0.046	0.044	0.055	0.056	0.057	0.124	0.125	0.153
100	0.048	0.048	0.047	0.060	0.059	0.062	0.142	0.142	0.180
250	0.049	0.050	0.049	0.062	0.062	0.066	0.156	0.155	0.204
500	0.048	0.050	0.048	0.063	0.063	0.068	0.160	0.161	0.215
1000	0.051	0.049	0.051	0.064	0.063	0.067	0.163	0.166	0.220

Table 3.11 Error variances generated as $\omega_{iit} = 1 + \frac{t-1}{T-1}9$, $t = 1, \ldots, T$, $\rho = 0.0$. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test. At the sample sizes 25, 50, 100, 250, 500 and 1000, the estimated rejection frequencies of the test of constant variance (3.13) are, when testing against:

	White, $n = 4$			ARCH(1)			Smooth transition		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.143	0.140	0.173	0.060	0.059	0.070	0.341	0.340	0.459
50	0.163	0.165	0.213	0.111	0.110	0.145	0.671	0.670	0.865
100	0.176	0.176	0.237	0.180	0.180	0.248	0.943	0.943	0.996
250	0.184	0.184	0.253	0.331	0.331	0.476	1.000	1.000	1.000
500	0.189	0.188	0.262	0.533	0.530	0.731	1.000	1.000	1.000
1000	0.189	0.190	0.268	0.791	0.789	0.945	1.000	1.000	1.000

Table 3.12 Error variances generated as $\omega_{iit} = 1 + 2I\left(t > \left[\frac{T}{2}\right]\right)$, $\rho = 0.0$, $t = 1, \ldots, T$. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test. At the sample sizes 25, 50, 100, 250, 500 and 1000, the estimated rejection frequencies of the test of constant variance (3.13) are, when testing against:

	White, $n = 4$			ARCH(5)			Smooth transition		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.095	0.097	0.105	0.045	0.044	0.046	0.110	0.108	0.121
50	0.116	0.115	0.139	0.085	0.085	0.096	0.254	0.254	0.343
100	0.122	0.120	0.148	0.149	0.149	0.192	0.508	0.511	0.703
250	0.119	0.119	0.150	0.283	0.281	0.401	0.904	0.905	0.989
500	0.117	0.117	0.148	0.460	0.461	0.653	0.997	0.997	1.000
1000	0.115	0.114	0.144	0.721	0.723	0.911	1.000	1.000	1.000

Table 3.13 Error variances generated as $\omega_{iit} \sim U(1, 10)$, $t = 1, \ldots, T$, $\rho = 0.0$. 'eq. 1' and 'eq. 2' corresponds to the univariate tests of each equation, and 'joint' corresponds to the bivariate test. At the sample sizes 25, 50, 100, 250, 500 and 1000, the estimated rejection frequencies of the test of constant variance (3.13) are, when testing against:

	White, $n = 4$			ARCH(5)			Smooth transition		
	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint	eq. 1	eq. 2	joint
25	0.131	0.133	0.160	0.026	0.026	0.025	0.065	0.064	0.067
50	0.156	0.156	0.202	0.048	0.048	0.052	0.077	0.077	0.085
100	0.168	0.169	0.224	0.074	0.077	0.091	0.082	0.081	0.093
250	0.176	0.179	0.244	0.103	0.104	0.132	0.086	0.087	0.101
500	0.182	0.183	0.254	0.118	0.118	0.153	0.088	0.087	0.102
1000	0.185	0.181	0.257	0.128	0.129	0.169	0.088	0.088	0.104

Bibliography

[1] Bollerslev, T., (1990), "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model", *Review of Economics and Statistics*, 72, 498-505.

[2] Box, G. E. P., and Jenkins, G. M., (1970), *Time series analysis: forecasting and control*, San Fransisco, Holden Day.

[3] Breusch, T. S., and Pagan, A. R., (1979), "A simple test for heteroscedasticity and random coefficient variation", *Econometrica*, 47, 1287-1294.

[4] Cecchetti, S. G., Cumby, R. E., and Figlewski, S., (1988), "Estimation of the optimal futures hedge", *Review of Economics and Statistics*, 70, 623-630.

[5] Engle, R. F., (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", Econometrica, 50, 987-1007.

[6] Engle, R. F., (2002), "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models", *Journal of Business and Economic Statistics*, 20, 339-350.

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

[8] Lin, C.-F. J., and Teräsvirta, T., (1994), "Testing the constancy of regression parameters against continuous structural change", *Journal of Econometrics*, 62, 211-228.

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

[10] Medeiros, M. C., and Veiga, A., (2003), "Diagnostic checking in a flexible nonlinear time series model", *Journal of Time Series Analysis*, (forthcoming).

[11] Tse, Y. K., (2000), "A test for constant correlations in a multivariate GARCH model", *Journal of Econometrics*, 98, 107-127.

[12] Tse, Y. K., and Tsui, A. K. C., (2002), "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations", *Journal of Business and Economic Statistics*, 20, 351-362.

[13] White, H., (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 48, 817-838.

Chapter 4

Estimating confidence regions over bounded domains

4.1 Introduction

Nonparametric density estimation by kernel techniques is a standard statistical tool in the estimation of a density function in situations where its parametric form is assumed unknown. This paper has to do with a special situation where the observations used for estimating the density are obtained from an unknown member of a known family of models. Information about the properties of the model may then be used in estimating the density function of interest. It may be a joint density of some random variables that in turn are functions of observations generated by a model belonging to a known family of models.

The example motivating this paper is one in which the problem is to estimate confidence regions for the combination of kurtosis and first-order autocorrelation of squared observations of a generalized autoregressive heteroskedasticity (GARCH) process. Such confidence regions can be obtained nonparametrically by first estimating the joint density of the kurtosis and autocorrelation estimators from a random sample. The interest in this problem originates from considerations in Teräsvirta (1996) who discussed the ability of some simple volatility models to reproduce so-called stylized facts evident in financial return series. In that case, the joint density of the two estimators has a form such that standard density estimation techniques based on a linear grid do not work properly. They have to be modified and such a modification is the topic of this paper.

The plan of the paper is as follows. Density estimation in general terms is discussed in Section 2. The GARCH(1, 1) model is presented and the choice of grid discussed in Section 3. In Section 4 the leading example, estimation of confidence regions to the kurtosis and first-order autocorrelation pair implied by a GARCH(1, 1) model, is considered. Concluding remarks are found in Section 5.

4.2 Kernel estimation

Let X_t, $t = 1, \ldots, T$, be a random sample from an n-dimensional probability distribution whose density $f(x)$ is to be estimated. The kernel density estimator with kernel K and window width h, also called the smoothing parameter or bandwidth in the literature, is defined in its most simple form by

$$\widehat{f}(x) = \frac{1}{h^n T} \sum_{t=1}^{T} K\left\{\frac{1}{h}(x - X_t)\right\} \tag{4.1}$$

where the kernel function $K(x)$ is a function that satisfies

$$\int_{\mathbb{R}^n} K(x)\,dx = 1. \tag{4.2}$$

This paper will focus on estimating confidence regions in two dimensions, so that in the remainder of the paper we assume that $n = 2$. Generalizing the method to higher dimensions is, however, conceptually straightforward.

Estimating the density function of a stochastic variable using the kernel estimator can in a simple way be described as approximating it by a sum of 'bumps' centered at the observations. The kernel function K determines the shape of the bumps while the window width h determines their width. When it comes to choosing kernel function K and window width h several possibilities exist, see Silverman (1986) and Scott (1992) for details. The kernel function is usually a radially unimodal probability density function. The most common ones are the Epanechnikov kernel

$$K_e(x) = \begin{cases} \frac{2}{\pi}(1 - x'x) & \text{if } x'x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

and the standard multivariate normal density function

$$K_n(x) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}x'x\right). \tag{4.4}$$

Two other kernels with some advantages over the Epanechnikov and normal kernel have the form

$$K_2(x) = \begin{cases} \frac{3}{\pi}(1 - x'x)^2 & \text{if } x'x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

$$K_3(x) = \begin{cases} \frac{4}{\pi}(1 - x'x)^3 & \text{if } x'x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

Kernels K_2 and K_3 have higher differentiability properties than the Epanechnikov kernel. These are properties that the resulting density estimate also will have. In addition, they can be calculated more quickly than the normal kernel (4.4), which saves valuable estimation time.

When estimating the density, it is a good idea to use a kernel function that has the same shape as the data cloud itself. If that is not possible the data can be 'pre-whitened' by a linear transformation such that it will have a unit covariance matrix. A density function is then estimated for the transformed data, which finally is transformed back to represent the density of the original

variable. In the two-dimensional case, this is equivalent to using a density estimate of the form

$$\widehat{f}(x) = \frac{|S|^{-1/2}}{h^2 T} \sum_{t=1}^{T} k\left\{ \frac{1}{h^2} (x - X_t)' S^{-1} (x - X_t) \right\} \tag{4.7}$$

where the function k is given by $k(x'x) = K(x)$, S is the sample covariance matrix of the data and h the window width.

It turns out, in practice, that the choice of the functional form of the kernel is not crucial, whereas the choice of h is very important. The window width h implies that the kernel placed on each observation is scaled equally in all directions. If h is chosen to be too small, too few observations will have a non-negligible weight and the resulting density estimate will appear rough or undersmoothed. If it is chosen too large, too many observations will have a non-negligible weight, which leads to oversmoothing the density and suppressing some of its characteristic features. In some cases it would be more appropriate to use several values of h, even a separate one for each data point if necessary. This is motivated if, for example, the observations are spreading out much more in one of the coordinate directions than in the other. Furthermore, obtaining a good estimate of the tail, or outer region, of the density can be a problem. Few observations in the tail can give a jumpy or edgy estimate since the given h would undersmooth the tail. For the purposes of this paper such modifications may not be necessary, however. When only confidence regions are being considered, the tail problem can be largely ignored.

4.3 Confidence regions for kurtosis and first-order autocorrelation of squared observations in a GARCH(1, 1) model

The density estimation technique discussed in this paper does not rely on an existing one-to-one data transformation as the one considered in the previous section and exemplified by equation (4.7). There may be situations where finding a reasonable transformation may not be easy but where other information is available. By that I mean information that is useful in defining the grid that determines the values x where the density is approximated. Normally, this grid is linear and consists of equidistant points in the original coordinate system, but in some cases other solutions may be available. In order to illustrate such a situation, consider the standard GARCH(1, 1) model of Bollerslev

(1986),

$$\begin{cases} y_t = \mu + \varepsilon_t \\ \varepsilon_t = z_t h_t \\ h_t^2 = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}^2, \end{cases} \tag{4.8}$$

where h_t^2, the conditional variance of ε_t (and y_t), is a positive-valued function, which implies that the parameters α_0, α, $\beta > 0$, and it is assumed that $z_t \sim nid(0,1)$, $t = 1, \ldots, T$. Furthermore, it is assumed that $E\varepsilon_t^4 < \infty$. A necessary and sufficient condition for this is that $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$ in (4.8). When this restriction holds, the kurtosis of ε_t and the first-order autocorrelation of ε_t^2 in (4.8) are given by the equations:

$$K_4 = \frac{3(1-(\beta+\alpha)^2)}{1-\beta^2-2\alpha\beta-3\alpha^2} \tag{4.9}$$

$$\rho_1 = \frac{\alpha(1-\beta^2-\alpha\beta)}{1-\beta^2-2\alpha\beta}, \tag{4.10}$$

see Bollerslev (1986, 1988) and He and Teräsvirta (1999).

The problem is the following. Suppose we estimate the parameters of model (4.8) from a time series. We can then estimate the kurtosis and the first-order autocorrelation by plugging in the maximum likelihood estimates for α and β into definitions (4.9) and (4.10). Call the resulting pair of values $\left(\widehat{K}_4, \widehat{\rho}_1\right)$. The question of interest is how to obtain confidence regions for the true kurtosis/autocorrelation combination. In principle, this can be done by simulation. One draws N pairs $\left(\alpha_{(i)}, \beta_{(i)}\right)$ from the asymptotic distribution of the maximum likelihood estimators $\widehat{\alpha}$ and $\widehat{\beta}$, which under regularity conditions including $E\left|\varepsilon_t^2\right|^{2+\delta} < \infty$, $\delta > 0$, see Berkes, Horváth and Kokoszka (2003), is normal. One then computes the values of (4.9) and (4.10) for each pair, called $\left(K_{4(i)}, \rho_{(i)}\right)$ where $i = 1, \ldots, N$. The pairs $\left(\alpha_{(i)}, \beta_{(i)}\right)$ for which $\beta_{(i)}^2 + 2\alpha_{(i)}\beta_{(i)} + 3\alpha_{(i)}^2 \geq 1$ are discarded because for them the fourth moment $E\varepsilon_t^4$ does not exist and (4.9) and (4.10) are therefore not defined. The confidence regions can now be estimated from the set $\left\{\left(K_{4(i)}, \rho_{(i)}\right) : i = 1, \ldots, N\right\}$ using an appropriate kernel method.

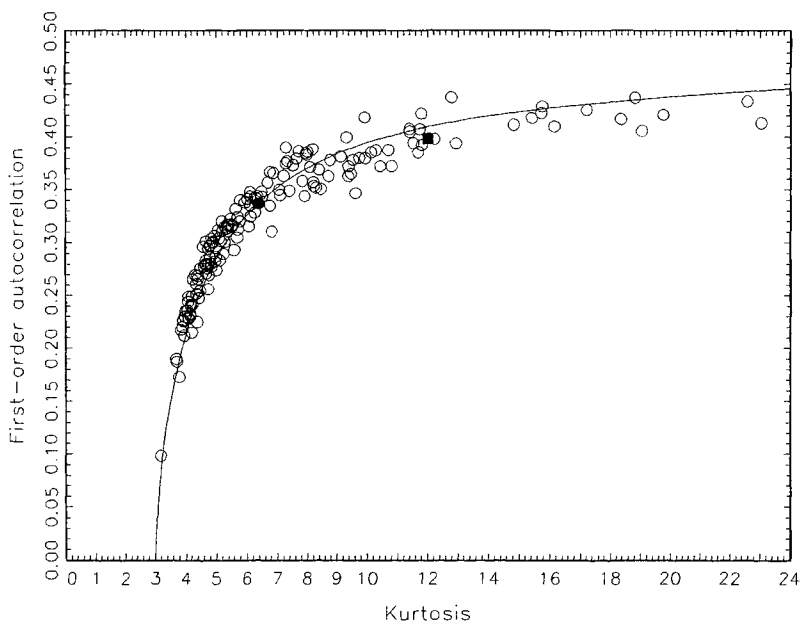
In order to illustrate the situation, consider the GARCH(1,1) model for which

$$h_t^2 = 0.05 + 0.19121\varepsilon_t^2 + 0.75879h_{t-1}^2. \tag{4.11}$$

The solid square in Figures 4.1 and 4.2 shows the values of the kurtosis of ε_t and the first-order autocorrelation of ε_t^2 of this model. After estimating its

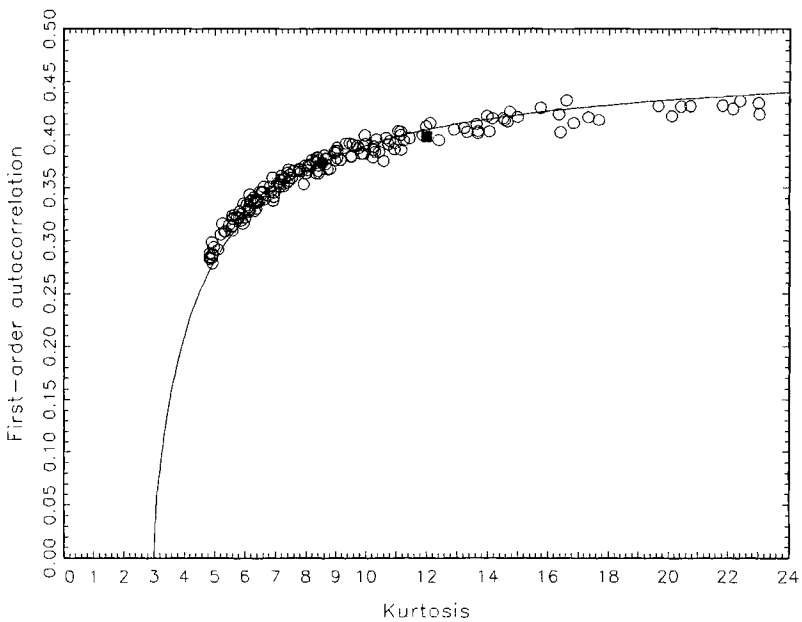
parameters, the kurtosis/autocorrelation pair can be estimated, and the solid circle in Figures 4.1 and 4.2 represents the estimated pair $\left(\widehat{K}_4, \widehat{\rho}_1\right)$. The empty circles represent the generated pairs $\left(K_{4(i)}, \rho_{1(i)}\right)$, $i = 1, \ldots, N$. Finally, the solid line is an isoquant consisting of all points for which $\alpha + \beta = c$, $c =$ constant, in the GARCH(1, 1) model. The constant has been selected such that $c = \widehat{\alpha} + \widehat{\beta}$. For a given value of c, increasing α, and at the same time decreasing β, is equivalent to moving on the isoquant to the right. This implies that for any $c > 0$, both the kurtosis and the autocorrelation are increasing functions of α.

Figure 4.1 Simulated kurtosis/autocorrelation combinations for the GARCH(1, 1) with $(\alpha_0, \alpha, \beta) = (0.05, 0.19121, 0.75879)$, and an isoquant $\alpha + \beta = 0.93868, T = 1000$, $N = 200$, ■ =true, ● =estimated, ⊙ =generated.



An interesting observation from Figures 4.1 and 4.2 is that the simulated points, while clustering close to $\left(\widehat{K}_4, \widehat{\rho}_1\right)$, are clustering around the isoquant defined by $\widehat{\alpha}$ and $\widehat{\beta}$, and spreading more along than around the isoquant. The point cloud consisting of the points $\left(K_{4(i)}, \rho_{1(i)}\right)$ thus has a very particular shape. Comparing Figures 4.1 (based on $T = 1000$) and 4.2 (based on $T = 5000$) it is also seen that with an increasing sample size, the simu-

Figure 4.2 Simulated kurtosis/autocorrelation combinations for the GARCH(1,1) with $(\alpha_0, \alpha, \beta) = (0.05, 0.19121, 0.75879)$, and an isoquant $\alpha + \beta = 0.94372, T = 5000$, $N = 200$, ■ =true, ●=estimated, ⊙ =generated.



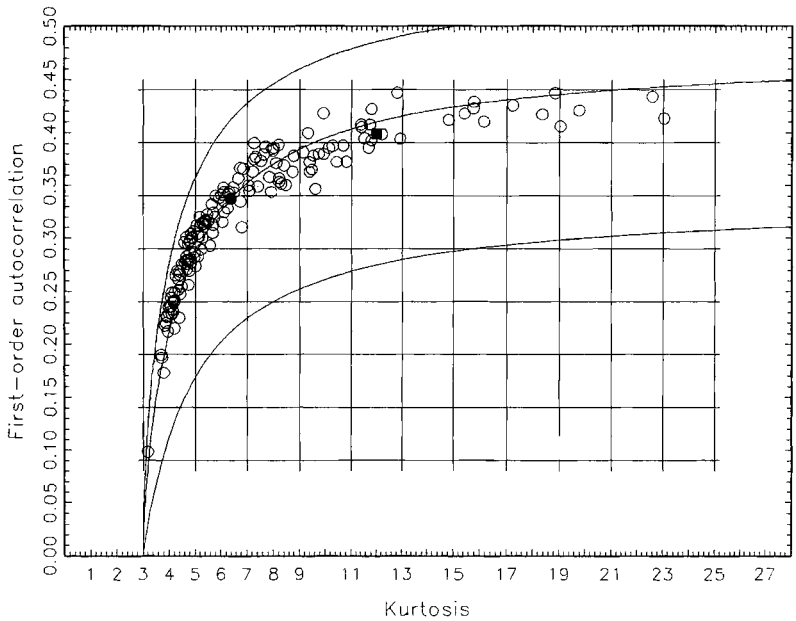
lated pairs become more concentrated around $\left(\widehat{K}_4, \widehat{\rho}_1\right)$, as can be expected. It may be worth noting that both the kurtosis and the autocorrelation are underestimated even when $T = 5000$.

4.3.1 Choosing the grid

The grid points x where $\widehat{f}(\cdot)$ is defined in (4.1) and (4.7) should cover the whole domain of interest. This means that they should cover all the points where the underlying density function is believed to be defined. In many cases the density is defined for all values $x \in \mathbb{R}^2$, in which case a linear grid is a natural choice. A linear grid usually consists of equidistant points between the smallest and the largest value in both dimensions. Figure 4.3 shows such a grid for the data set from Figure 4.1. The kernel estimation method will then assign a weight to each point x in the grid depending on the choice of kernel function and window width h. The resulting density estimate will, as a consequence, be defined for all values x in the grid with a nonzero weight.

Consider now Figure 4.3 where, except for the points and isoquant through

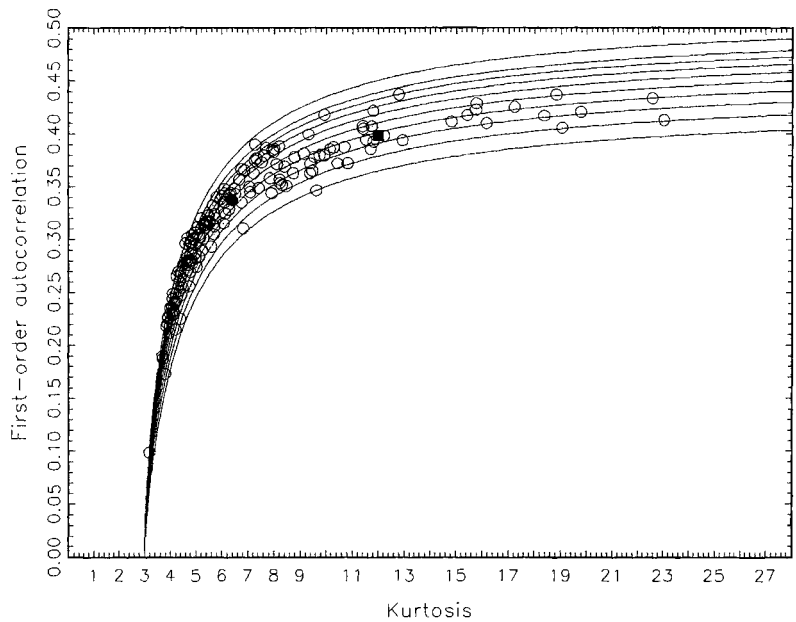
Figure 4.3 Linear grid for the data set and isoquant $\alpha + \beta = 0.93868$ and isoquants $\alpha+\beta = 0.8$ (highest isoquant) and $\alpha+\beta = 0.99999$ (lowest isoquant).



$\left(\widehat{K}_4, \widehat{\rho}_1\right)$, the isoquants for $\alpha+\beta = 0.8$ and $\alpha+\beta = 0.99999$ are included. Note that the necessary and sufficient condition for the second moment of ε_t to exist is $\alpha + \beta < 1$. Furthermore, by construction, the GARCH(1, 1) model implies a kurtosis of ε_t at least equal to three. Thus, no values of α and β in (4.9) and (4.10) would then result in a kurtosis/autocorrelation point in the lower right corner of the figure, or in points with a kurtosis less than three. The use of the linear grid for this data set is thus not a practical choice because such a grid would include a number of irrelevant points. This fact in turn would result in an unrealistic and unacceptable estimate of the density function. This is true even if the step size between the points in the grid is made smaller since, even for a very narrow window width h, nonzero weights will be given to grid points close to the boundary but outside the area of interest. The estimated density should equal zero for such points. The problem may be solved by a suitable reflection or replication technique as in Boneva, Kendall and Stefanov (1971), or some boundary kernel method for irregular regions as in Staniswalis, Messer and Finston (1990). Spline smoothing, as in Ramsay (2002), could also be applied, but none of these methods will be considered here.

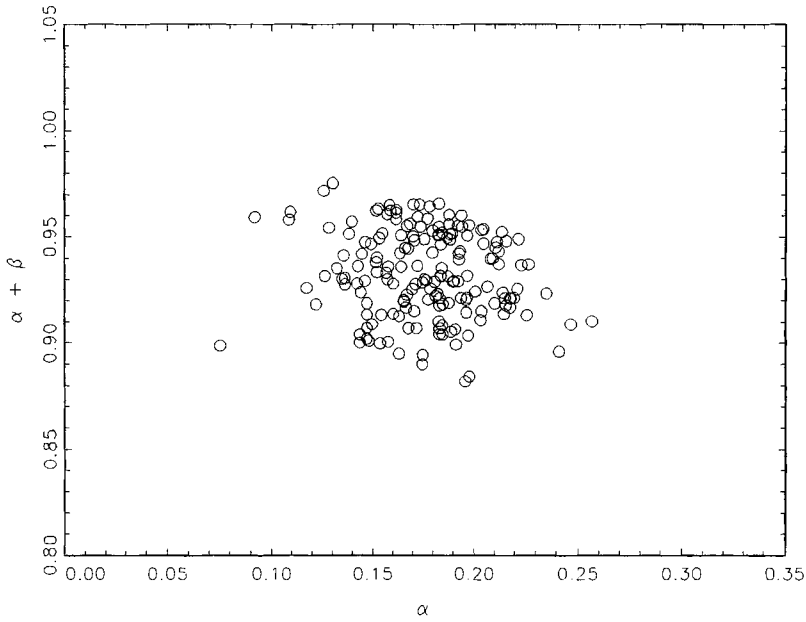
To obtain a realistic density function estimate it is thus necessary, before constructing the grid, to know the points for which the underlying density is defined. Since all points in Figure 4.3 lie on some isoquant, the 'definition domain' of the density function has the same shape as the isoquants. As a consequence, using isoquants as grid lines appears an attractive solution. The resulting nonlinear grid then clearly only covers the theoretically possible kurtosis/autocorrelation points, which solves the problem of assigning nonzero weights to points outside the domain of interest. Figure 4.4 shows the nonlinear grid for the data set in Figure 4.1 using ten isoquants from $\alpha + \beta = 0.88$ to $\alpha + \beta = 0.98$. In practice, both the number of isoquants in the grid and the number of points on the isoquants influence the resulting density estimate. Furthermore, in contrast with the linear grid, the nonlinear grid does not consist of equidistant points.

Figure 4.4 Nonlinear grid with ten isoquant curves from $\alpha + \beta = 0.88$ to $\alpha + \beta = 0.98$ for the data set in Figure 4.1.



As mentioned in the previous section, the data can in some cases be 'prewhitened' by a one-to-one transformation. Using the nonlinear grid when estimating the density function will in fact yield results similar to the ones from such a transformation. To illustrate this, consider Figure 4.5 which shows the generated pairs in the $(\alpha, \alpha + \beta)$ coordinate system. The value of $\alpha + \beta$

Figure 4.5 Representation of the data set in the $(\alpha, \alpha + \beta)$ coordinate system.



determines on which isoquant an observed point lies, whereas the value of α indicates where on the isoquant the point appears. When using the nonlinear grid, the kernel function is estimated for each isoquant, moving along the curve point by point. When all points on all isoquants have been considered, all values of α and $\alpha + \beta$ have also been considered as if the density function had been estimated on a linear grid over α and $\alpha + \beta$. Note that the data have not been actually transformed, the reason being that such a transformation may not be applicable. This is because it is not known whether or not the nonlinear mapping from $(\alpha, \alpha + \beta)$ to (K_4, ρ_1), or (α, β) to (K_4, ρ_1) in (4.9) and (4.10), has a unique inverse.

Note that, without any knowledge about an underlying model, estimating the confidence region to the data set in Figure 4.1 would be very difficult. In such a case, no information about for which pairs of kurtosis and autocorrelation the density is defined would exist. Using the nonlinear grid would then be incorrect, since it excludes points where the density could be defined. The density would thus have to be estimated for the data set as it is, or possible after some transformation of the data or modification of the Kernel method.

4.4 Estimating confidence regions

We will now illustrate the method by estimating approximative confidence regions for the true kurtosis/autocorrelation pair in Figure 4.1. The confidence regions are based on the estimated density function using the nonlinear grid and the criterion that the region should occupy the smallest possible volume in the sample space. As pointed out by Hyndman (1996) this criterion is equivalent to the requirement that every point in the region should have probability at least as large as every point outside the region. This requirement seems intuitive sensible and leads to regions called highest density regions (HDR). A formal definition in Hyndman (1996) is the following:

Definition 1 *Let $f(x)$ be the density function of a random variable X. Then the $100(1-\alpha)\%$ HDR is the subset $R(f_\alpha)$ of the sample space of X such that*

$$R(f_\alpha) = \{x : f(x) \geqslant f_\alpha\}$$

where f_α is the largest constant such that $\Pr(X \in R(f_\alpha)) \geqslant 1-\alpha$.

Figure 4.1 shows that the kurtosis and first-order autocorrelation are positively correlated. Using the weighted version of the density function estimate, equation (4.7), then results in a density function estimate aligned with the data cloud. This is equivalent to using a linear transformation in order to 'pre-whiten' the data. It will, however, not result in a unit covariance matrix of the data because the relationship between the kurtosis and the first-order autocorrelation is nonlinear, but it can be regarded as the 'best possible linear transformation' of the data set.

In practice, as mentioned in Section 2, the choice of kernel function is not important. The kernel K_2, defined in (4.5), is then chosen to save estimation time. Applying the kernel function K_3 in (4.6) results in only small differences in the estimated confidence region compared to the regions obtained using K_2. The choice of h, the smoothing parameter or bandwidth, requires more care. Since only the confidence regions are of interest it can in some cases be difficult to determine if the estimated density is undersmoothed or oversmoothed. A good strategy in practice is to plot the estimated density surface for a wide range of different values of h, both undersmoothing and oversmoothing the density. From this collection of density estimates one may subjectively choose the bandwidth yielding an acceptable smooth density estimate. In this case, however, it is not possible to plot the estimated density surface using standard computer program packages, since the nonlinear grid does not consist of

equidistant points. Instead, h can be chosen by an analogous method. Estimate a confidence region with a very high confidence level, for example 99% or 99.9%, for a sequence of different values of h as before. Then, in a similar fashion, the region with acceptable smooth properties, chosen subjectively from an appropriate collection of confidence regions, all with the same confidence level, determines the bandwidth.

Figures 4.6, 4.7 and 4.8 show data points and estimated approximative 50%, 60%, 70%, 80%, and 90% HDR's of the true kurtosis/autocorrelation pairs. Figure 4.6 is based on the data set in Figure 4.1. The data points in Figures 4.7 and 4.8 are generated by the same method as for the data in Figure 4.1, considering two other sets of parameters in the GARCH(1,1) model (4.8). In Figure 4.7 the data is generated for the parameters $(\alpha_0, \alpha, \beta) = (0.10, 0.15, 0.75)$, and the data in Figure 4.8 are generated for $(\alpha_0, \alpha, \beta) = (0.01, 0.09, 0.90)$.

Figure 4.6 Approximative 50%, 60%, 70%, 80%, and 90% HDR's of the true kurtosis/autocorrelation combination. Data generated from (4.8) with $(\alpha_0, \alpha, \beta) = (0.05, 0.19121, 0.75879)$, $T = 1000$, $N = 200$, with a window width $h = 0.7$.

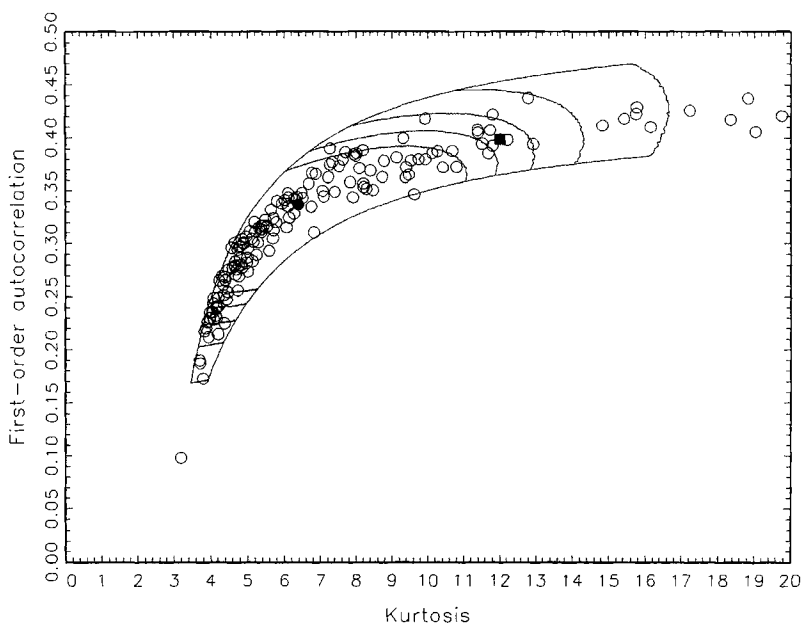
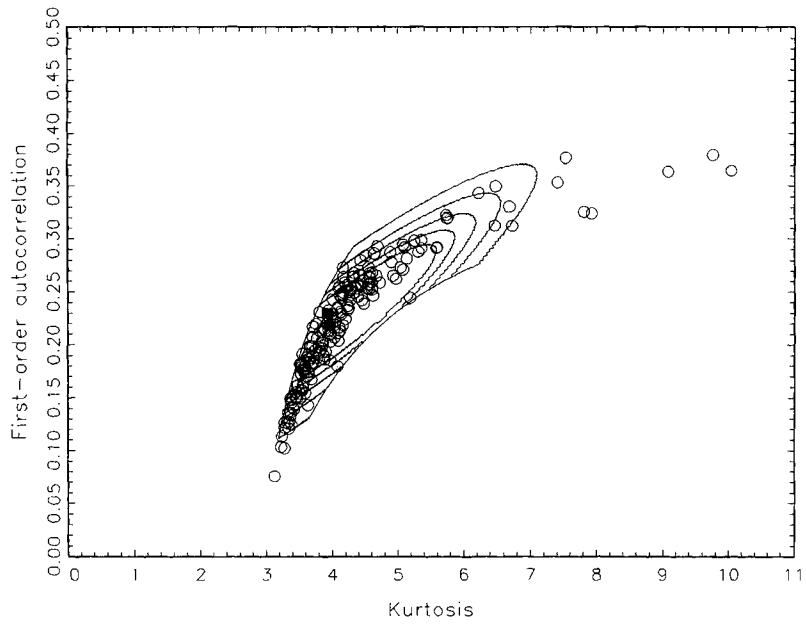


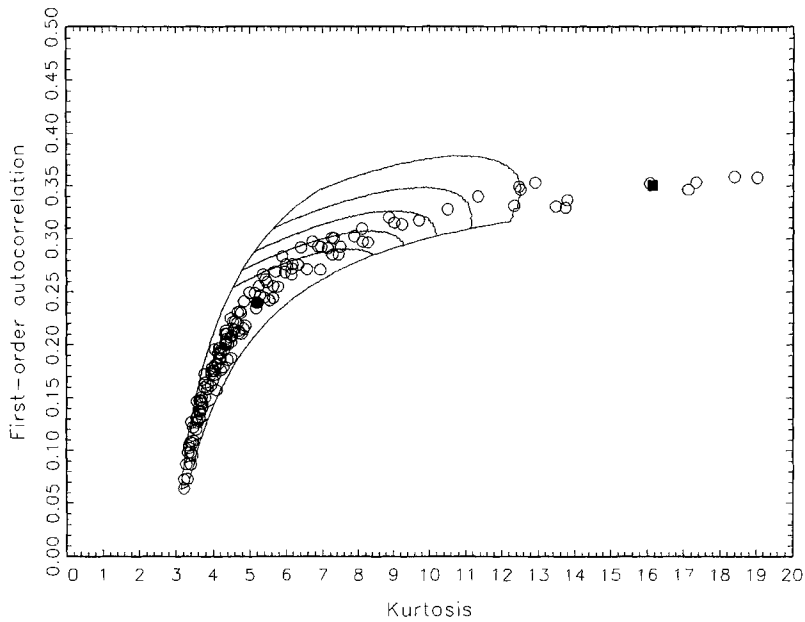
Figure 4.7 Approximative 50%, 60%, 70%, 80%, and 90% HDR's of the true kurtosis/autocorrelation combination. Data generated from (4.8) with $(\alpha_0, \alpha, \beta) = (0.10, 0.15, 0.75)$, $T = 1000$, $N = 200$, with a window width $h = 0.7$.



The solid circle and square in the figures are, as before, the estimated pair $\left(\widehat{K}_4, \widehat{\rho}_1\right)$ and the true kurtosis/autocorrelation pair, respectively. The density estimate can be improved by including more isoquants in the grid or estimating the kernel at more points on the isoquants. The latter is useful in particular when the kurtosis is high because the estimate of the tail of the density is only based on a small number of observations. The regions in Figures 4.6, 4.7 and 4.8 are estimated using 100 isoquants and considering at least 1000 points on each isoquant.

A well-known feature of the GARCH(1, 1) model is that as the sum $\alpha + \beta$ tends to 1, it becomes more difficult to obtain good estimates of the parameters α and β, which in turn indirectly affects the estimated kurtosis and autocorrelation. This feature can be observed in the figures. When the sum $\alpha + \beta = 0.90$ as in Figure 4.7, the points in the figure are more concentrated around the true kurtosis/autocorrelation point than is the case in Figures 4.6 and 4.8 where $\alpha + \beta = 0.95$ and $\alpha + \beta = 0.99$ respectively. This affects the

Figure 4.8 Approximative 50%, 60%, 70%, 80%, and 90% HDR's of the true kurtosis/autocorrelation combination. Data generated from (4.8) with $(\alpha_0, \alpha, \beta) = (0.01, 0.09, 0.90)$, $T = 1000$, $N = 200$, with a window width $h = 0.7$.



confidence regions directly, as the regions in Figures 4.7 and 4.8 are wider than the ones in Figure 4.6. The three data sets are of course only one realization of each data generating process, but the deviation of the estimated kurtosis/autocorrelation point from the true appears in general to increase with the sum $\alpha + \beta$. The 90% confidence region in Figure 4.8 does not even cover the true kurtosis/autocorrelation point.

4.5 Conclusions

This paper presents a method for estimating confidence regions over bounded domains if no one-to-one transformation of the considered data set exists, or if the existence of such a transformation is difficult to verify. The idea is to estimate approximate confidence regions by simple kernel estimation using a nonlinear grid implied by some parameter restrictions of an underlying model. Applying the method to three data sets, generated from the GARCH(1, 1) model, the resulting confidence regions cover a reasonable area of the def-

inition space, and are well aligned with the corresponding data sets. The method, as it has been presented here, can perhaps be improved on at least two points. One could choose an alternative kernel method which takes into account the nonlinear dependence in the data, or apply some form of adaptive kernel method using a varying window width.

Bibliography

[1] Berkes, I., Horváth, L., and Kokoszka, P., (2003), "GARCH processes: structure and estimation", *Bernoulli*, 9, 201-227.

[2] Bollerslev, T., (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

[3] Bollerslev, T., (1988), "On the correlation structure for the generalized autoregressive conditional heteroskedastic process", *Journal of Time Series Analysis*, 9, 121-131.

[4] Boneva, L. I., Kendall, D. G., and Stefanov, I., (1971), "Spline transformations: three new diagnostic aids for the statistical data-analyst (with Discussion)", *Journal of the Royal Statistical Society*, 33, 1-70.

[5] He, C., and Teräsvirta, T., (1999), "Properties of moments of a family of GARCH processes", *Journal of Econometrics*, 92, 173-192.

[6] Hyndman, R., (1996), "Computing and graphing highest density regions", *The American Statistician*, 50, 120-126.

[7] Ramsay, T., (2002), "Spline smoothing over difficult regions", *Journal of the Royal Statistical Society*, 64, 307-319.

[8] Scott, D. W., (1992), *Multivariate density estimation: Theory, practice, and visualization*, John Wiley.

[9] Silverman, B. W., (1986), *Density estimation for statistics and data analysis*, Monographs on statistics and applied probability, 26, Chapman & Hall.

[10] Staniswalis, J. G., Messer, K., and Finston, D. R., (1990), "Kernel estimators for multivariate smoothing", Technical report 90-01, Biostatistics, Virginia Commonwealth University.

[11] Teräsvirta, T., (1996), "Two stylized facts and the GARCH(1, 1) model", Stockholm School of Economics, *Working Paper Series in Econometrics and Finance*, No. 96.

EFI
The Economic Research Institute

Reports since 1998
A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title

2003
Andersson, H., Valuation and Hedging of Long-Term Asset-Linked Contracts.
Bergman, M., Essays on Human Capital and Wage Formation.
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.
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.
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 stadsdelsnämnds-reformen.
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 stadsdels-nämndsreformen.

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 budget-avvikelser.

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 stadsdels-nämndsreformen – 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 stadsdelsnämndsreformen – 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.

1998

Andersson, M., On Testing and Forecasting in Fractionally Integrated Time Series Models.
Berg-Suurwee, U., Styrning av kultur- och fritidsförvaltning innan stadsdelsnämndsreformen.
Berg-Suurwee, U., Nyckeltal avseende kultur- och fritidsförvaltning innan stadsdelsnämndsreformen.

Bergström, F., Essays on the Political Economy of Industrial Policy.
Bild, M., Valuation of Takeovers.
Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform – en analys av Stockholmsmodellen.
Eriksson-Skoog, G., The Soft Budget Constraint: The Emergence, Persistence and Logic of an Institution. The Case of Tanzania 1967-1992.
Gredenhoff, M., Bootstrap Inference in Time Series Econometrics.
Ioannidis, D., I nationens tjänst? Strategisk handling i politisk miljö – en nationell teleoperatörs interorganisatoriska, strategiska utveckling.
Johansson, S., Savings Investment, and Economic Reforms in Developing Countries.
Levin, J., Essays in Company Valuation.
Ljunggren, U., Styrning av grundskolan i Stockholms stad innan stadsdelsnämndsreformen.
Mattsson, S., Från stat till marknad – effekter på nätverksrelationer vid en bolagiseringsreform.
Nyberg, A., Innovation in Distribution Channels – An Evolutionary Approach.
Olsson, P., Studies in Company Valuation.
Reneby, J., Pricing Corporate Securities.
Roszbach, K., Essays on Banking Credit and Interest Rates.
Runsten, M., The Association Between Accounting Information and Stock Prices. Model development and empirical tests based on Swedish Data.
Segendorff, B., Essays on Bargaining and Delegation.
Sjöberg, L., Bagozzi, R., Ingvar, D.H., Will and Economic Behavior.
Sjögren, A., Perspectives on Human Capital: Economic Growth, Occupational Choice and Intergenerational Mobility.
Studier i kostnadsintäktsanalys. Jennergren, P., (red)
Söderholm, J., Målstyrning av decentraliserade organisationer. Styrning mot finansiella och icke-finansiella mål.
Thorburn, K., Cash Auction Bankruptcy and Corporate Restructuring
Wijkström, F., Different Faces of Civil Society.
Zethraeus, N., Essays on Economic Evaluation in Health Care. Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations.