Three Essays on Unit Roots and Nonlinear Co-Integrated Processes

Inaugural–Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts– und Gesellschaftswissenschaften
durch die
Rechts– und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von
Jürgen Gaul
aus Olching

Bonn 2008

Dekan:	Prof. Dr. Erik Theissen
Erstreferent:	Prof. Dr. Jörg Breitung
Zweitreferent:	Prof. Dr. Erik Theissen

Tag der mündlichen Prüfung: 7. Oktober 2008

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

to my family

ACKNOWLEDGEMENTS

While I was writing this thesis, I received help and support of many people and institutes.

First of all, I would like to thank my principal supervisor Jörg Breitung. When being a student at Ludwig-Maximilians-Universität München, he awaked my interest in the very exciting area of integrated and co-integrated processes. During the past years, he gave me the freedom to develop my own research ideas and provided excellent guidance whenever I needed. In many discussions, I benefitted a lot from his comments and suggestions to improve my work.

I am very thankful to my co-author Erik Theissen for the excellent co-operation. Furthermore, he accepted to act as referee on my thesis committee.

In my first year I took much advantage of a reading group on the book "Likelihood-based inference in cointgrated vector autoregressive models" by Søren Johansen. I would like to thank the participants Jörg Breitung, Norbert Christopeit, Oliver Fries, Simon Junker, Michael Massmann and Marc Schiffbauer for many fruitful discussions.

I am very indebted to Norbert Christopeit and Michael Massmann for their advice to tackle several econometric problems.

Next, I would like to thank Urs Schweizer, Georg Nöldeke and Jürgen von Hagen for managing the Bonn Graduate School of Economics and the DFG-Graduiertenkolleg "Quantitative Economics". I also want to thank the German Research Foundation (DFG) as well as the DAAD (Grant No. D/06/43977) for providing me financial support during my studies at the Bonn Graduate School of Economics and Yale University.

Parts of the research for this thesis were carried out at Yale University in New

Haven, CT, USA. I am indebted to Don Andrews, Brendan Beare, Taisuke Otsu and Peter Phillips for several constructive discussions.

I want to thank my fellow graduate students Johannes Abeler, Steffen Altmann, Eva Benz, Michael Brei, Almira Buzaushina, Meral Cakici, An Chen, Zeno Enders, Michael Evers, Oliver Fries, Bernhard Ganglmeier, Sotiris Georganas, Simon Junker, Jördis Klar, Lars Koch, Julia Nafziger, Michal Paluch, Marc Schiffbauer, Bernd Schlusche, David Schröder, Klaas Schulze, Markus Sonntag, Xia Su, Jörn Tenhofen, Matthias Wibral and Michael Zähringer who all contributed in their own way that I had such a great time in Bonn. I will never forget the time we have spent together, the trip to Barcelona and the innumerable parties.

I would like to give my gratitude to my friends Marcus Trügler, Christine Bögl, Michael Kettner, Matthias Lux, Johannes Richter, Stefan Täumer and Simon Zehnder for cheering me up in bad times.

Most of all, I would like to thank my parents, my sister Andrea and her husband Florian, and my brother Christoph and his wife Kristy. I thank them for their endless encouragement, patience and imperturbable belief in me. Words fail to express my gratitude.

Thank you all very much!

Contents

List of Tables

List of Figures

Introduction

Many macroeconomic and financial key variables such as e.g. consumption, investment, gross domestic product and interest rates, display non-stationary features such as trends or changing variances. A non-stationary stochastic process that can be made stationary by taking first differences, is called *integrated of order one (I(1))*. If the variables in a p-dimensional vector $y_t = (y_{1,t}, \ldots, y_{p,t})^\top$, $t = 1, \ldots, T$, are I(1), then there may exist a linear combination of the variables such that the resulting process is stationary. Integrated processes with this property are said to be *cointegrated*. The concept of *cointegration*, introduced by Granger (1981) and Engle and Granger (1987), allows to describe equilibrium relationships between economic variables and, hence, bridges the gap between time series analysis and economics. For this reason, cointegration has become a popular tool for applied econometric work, e.g. Johansen and Juselius (1992).

In the last 25 years, both integrated and cointegrated processes have attracted a lot of attention in theoretical and applied time series econometrics. The seminal contributions by Dickey and Fuller (1979), Engle and Granger (1987), Phillips and Perron (1988) and Johansen (1988, 1991) provided a solid basis for numerous extensions of this field of research.

This dissertation sheds light on two important extensions of the unit root model and the linear vector error correction model (VECM). In the first chapter, I extend several state-of-the-art unit root tests in the presence of permanent variance changes and compare their finite sample behavior in an extensive simulation study. In the two remaining chapters, I concentrate on error correction models that allow for a nonlinear adjustment process. The second chapter is devoted to the statistical inference of a general three regime threshold VECM. In chapter three, I explore the dynamics of spot and future prices using a novel nonlinear error correction model.

In the next paragraphs, I present the main topics covered in this thesis in more detail.

The first chapter[1], is devoted to testing for a unit root under time-varying variances. Recent literature find that many macroeconomic and financial variables exhibit a change in unconditional volatility. The presence of non-stationary volatility alters the asymptotic distribution of unit root tests and may lead to a high rejection rate of the null hypothesis.

The aim of this chapter is to investigate the finite-sample properties of three recent unit root tests. First, Beare (2006) suggests to scale the innovations by a nonparametric estimator of their variances such that the first differences of the series are approximately homoskedastic. Then, the Phillips-Perron test is applied to the scaled data. Additionally, I apply the local-to-unity approach by Elliott et al. (1996) to the test by Beare (2006) in order to de-trend the series. Second, Cavaliere and Taylor (2008) propose to apply the wild bootstrap method to the M class of unit root tests. Third, I extend the ML coefficient test by Boswijk (2005) to allow for deterministic components.

The key results of the simulation study are as follows. First, the proposed detrended version of the Phillips-Perron test by Beare (2006) improves the power substantially. Second, empirical size and power of unit root tests requiring a nonparametric estimator of the variance path, depend considerably on the choice of the bandwidth parameter. Third, the ML test outperforms the remaining tests in terms of power.

In the second chapter[2] of this thesis, I focus on a three-regime threshold VECM whose dynamics are characterized by a piecewise linear VECM. The regime switches

[1]This chapter is based on Gaul (2008).

[2]This chapter is based on Gaul (2007).

depend on the magnitude of a stationary threshold variable crossing unknown threshold parameters. In contrast to previous literature, the model does not impose any rank restrictions on the long-run impact matrix.

To estimate the parameters of the model, I employ the constrained least squares method. I show that the estimator for the threshold parameters is consistent. Then, I introduce an information-based selection criterion to estimate the cointegration ranks. If the regimes are characterized by the *same* rank configuration, I construct an LM test to detect threshold effects. Since the threshold parameters are not identified under the null hypothesis, I show that the asymptotic distribution is non-standard and, in particular, depends on moments of the data set. A parametric bootstrap algorithm is proposed to simulate critical values.

An empirical application to the term structure of US interest rates is conducted to highlight the approach. The results confirm the intuition that the series are not cointegrated in the case of small deviations from the long-run equilibrium, but they become cointegrated for large deviations. Furthermore, the model clearly outperforms the random walk model and the linear error correction model in terms of forecast ability.

In chapter 3[3], I consider the dynamics of spot and futures prices in the presence of arbitrage. The cost-of-carry relation establishes the relationship between prices in spot and futures markets. In a frictionless world any deviations from this relation would be eliminated by arbitrage. In practice, however, the prices in both markets may and do differ for several reasons. First, the existence of transactions costs makes it unprofitable to exploit small deviations. Second, traders with access to private information may prefer to trade in a specific market. Consequently, prices in this market may reflect information earlier than prices in the other market.

I use a novel partially linear VECM where the adjustment process is modelled by a non-parametric function. Estimation of the model is non-standard and employs kernel

[3]This chapter is based on Gaul and Theissen (2007).

methods. The short-run dynamics are estimated by density-weighted OLS, whereas the adjustment process is estimated by a Nadaraya-Watson estimator. To detect whether the adjustment process is indeed nonlinear, I perform a non-parametric test whose test statistic is asymptotically normally distributed.

The model is implemented using data on the DAX index and the DAX futures contract. I find that the conjecture of a nonlinear adjustment is strongly supported by the non-parametric test. The speed of price adjustment is increasing almost monotonically with the magnitude of the price difference. The estimation results indicate that the futures market leads the spot market. Furthermore, I observe that the partially linear VECM clearly improves the forecasting ability compared with that of the linear VECM.

The next three chapters each present one idea as a self-contained unit.

Chapter 1

Unit Root Tests with Time-Varying Variances: A Simulation Study

1.1 Introduction

Many key variables such as gross domestic products, interest rates and exchange rates display non-stationary behavior. Since the seminal contributions by Dickey and Fuller (1979) and Phillips and Perron (1988) testing for a unit root has become a widely used tool in macroeconomics and finance. However, the derivation of the asymptotic distribution of unit root tests relies on the assumption of a constant variance of the innovations. For example, models of heteroskedasticity involving a sudden change in the volatility are very common in econometrics. Among others, Sensier and van Dick (2004) find that about 80% of 214 US macroeconomic time series feature a change in unconditional volatility.

To the author's best knowledge, Hamori and Tokihisa (1997) were the first to show that a single break in the variance can have a great impact on the asymptotic distribution of standard unit root tests. For this specific case, Kim, Leybourne and Newbold (2002) suggest a modified version of the Perron (1989, 1990) unit root tests. Cavaliere

(2004) develops a new asymptotic theory that allows for a wide class of heteroskedasticity. Based on the latter contribution, different approaches for testing for a unit root under heteroskedasticity has been proposed. In this chapter we investigate the finite sample properties of three state-of-the-art unit root tests.

Beare (2006) proposes a modification to the Phillips-Perron test that scales the innovations by a nonparametric estimator of their variances such that the first differences of the series are approximately homoskedastic. Then, the Phillips-Perron test is applied to the scaled data. The main advantage of this method is that the test statistic has the standard asymptotic distribution and hence the use of simulation methods is not necessary. Inspired by Elliott et el. (1996) we suggest a detrended variant of the test.

Recently, Cavaliere and Taylor (2008) analyze in the case of time-varying variances the asymptotic distribution of the M class of unit root tests originally introduced by Stock (1999). They show that the asymptotic distribution depends on the volatility path and therefore general critical values are not available. They establish that the wild bootstrap method is valid since it replicates the heteroskedasticity present in the original shocks in the resampled data.

Boswijk (2005) derives the asymptotic power envelope of unit root tests under heteroskedasticity when the series has no deterministic component. Furthermore, he constructs a class of feasible test statistics whose power functions are tangent to the power envelope at one point. Boswijk (2005) observes that the power of the Gaussian ML coefficient test is close to the power envelope. In this chapter we allow for deterministic components and propose to detrend the series using the local-to-unity approach by Elliott et al. (1996). Then, we perform the Gaussian ML coefficient test following Boswijk (2005).

The key results of the simulation study can be summarized as follows. First, the standard unit root tests exhibit serious size distortions for several models of heteroskedasticity. Second, the proposed detrended version of the Phillips-Perron test by Beare (2006) increases the power substantially. Third, empirical size and power of unit root tests requiring a nonparametric estimator of the variance path depend considerably on the choice of the bandwidth parameter. Fourth, the ML test outperforms the remaining tests in terms of power.

The remaining chapter is organized as follows. In section 1.2 we introduce the model and the basic assumptions. The test by Beare (2006) and its detrended version are outlined in section 1.3. Section 1.4 is devoted to the bootstrap M tests by Cavaliere and Taylor (2008). We describe the modification to the ML test by Boswijk (2005) in section 1.5. The simulation results for two different data generating processes are reported in section 1.6. Finally, we conclude.

1.2 Model

As in Beare (2006) and Cavaliere and Taylor (2008) we suppose that the data generating process is given by the econometric model

$$y_t = \delta^\top z_t + x_t, \qquad t = 0, \ldots, T \tag{1.1}$$

$$x_t = \rho x_{t-1} + g\left(\frac{t}{T}\right)\epsilon_t. \tag{1.2}$$

In (1.1), the deterministic components of the process are collected in z_t. Here, we restrict our attention to the leading cases of a constant, i.e. $z_t = 1$ and $\delta = \delta_0$, and a linear trend, i.e. $z_t = (1, t)^\top$ and $\delta = (\delta_0, \delta_1)^\top$.

The class of functions triggering heteroskedasticity is defined by the following assumption.

Assumption V: *The function* $g : [0,1] \to \mathbb{R}$ *is twice continuously differentiable except at a finite number of points. Furthermore, there exists a constant* $M \in (1, \infty)$ *such that* $M^{-1} < g(r) < M$ *for all* $r \in [0,1]$, $|g'(r)| < M$ *at which* $g'(r)$ *exists, and* $|g''(r)| < M$ *at which* $g''(r)$ *exists. Additionally, we assume that* g *is right-continuous.*

It encompasses several different forms of models including non-continuous functions with a finite number of breaks and trending functions. However, stochastic functions such as GARCH models, near-integrated and integrated models are not covered by our framework.

Furthermore, we suppose that the error process $(\epsilon_t)_{t=1\ldots,T}$ satisfies the following condition.

Assumption E: *The error process* $(\epsilon_t)_{t=1\ldots,T}$ *is a strictly stationary* α*-mixing process with* $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = 1$, $E|\epsilon_t|^p < \infty$ *for some* $p > 4$ *and with mixing coefficients* $(\alpha_m)_{m\geq 0}$ *satisfying* $\sum_{m=0}^{\infty} \alpha_m^{2(1/r-1/p)} < \infty$ *for some* $r \in (2, 4]$. *Additionally, we require that the long-run variance* $\sigma^2 = \sum_{j=-\infty}^{\infty} E(\epsilon_t \epsilon_{t+j})$ *is strictly positive and finite.*

We aim to test the null hypothesis $\rho = 1$ in the situation that the innovations may be heteroskedastic.

The following lemma by Cavaliere (2004) plays a central role in order to analyze the effect of heteroskedasticity on standard unit root tests. It extends the standard functional central limit theorem to the heteroskedastic case.

Lemma 1.1:

Suppose Assumption V and E hold. Let $u_t = g\left(\frac{t}{T}\right)\epsilon_t$. As $T \to \infty$,

$$S_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} u_t \Rightarrow \int_0^r g(s) dW(s) = W_g(r),$$

where $W(\cdot)$ denotes a standard Brownian motion.

In the homoskedastic case $g(s) = \sigma$ it follows that

$$S_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} \sigma \epsilon_t \Rightarrow \sigma \int_0^r dW(s) = \sigma W(r)$$

and, hence, Lemma 1.1 coincides with the standard FCLT.

Let $\hat{\rho}$ denote the least square estimator in (1.1) and (1.2) and set for simplicity $z_t = 0$. As shown by Cavaliere (2004), the asymptotic distribution of the standard Dickey-Fuller coefficient test is given by

$$T\hat{\rho} \Rightarrow \frac{\int_0^1 W_g(r)g(r)dW(r)}{\int_0^1 W_g(r)^2 dr}. \tag{1.3}$$

Note that the distribution on the right-hand side in (1.3) differs substantially from the standard Dickey-Fuller distribution unless $g(s) = \sigma$. Therefore, we expect that standard unit root tests fail to work under heteroskedasticity. In the subsequent sections we describe recent adjustments to this lack of robustness, evaluate the finite sample properties of the tests and compare them with the standard tests.

1.3 The Modified Phillips-Perron Test

1.3.1 Phillips-Perron test by Beare (2006)

Recently, Beare (2006) has developed a correction to the standard Phillips-Perron test that allows for a very general class of heteroskedasticity. As shown by Beare (2006), the approach offers the advantage that the test statistic possesses the standard Dickey-Fuller distribution. The test statistic can be derived as follows. Under the null hypothesis, $\rho = 1$, it follows from (1.1) and (1.2) that

$$\Delta y_t = \delta_1 + g\left(\frac{t}{T}\right)\epsilon_t.$$

Regressing Δy_t on a constant, we obtain the least square estimator $\hat{\delta}_1 = \frac{1}{T}\sum_{t=1}^T \Delta y_t$. A nonparametric estimator of g is given by the Nadaraya-Watson estimator

$$\hat{g}(r) = \left[\frac{\sum_{t=1}^T K\left(\frac{t/T-r}{h}\right)(\Delta y_t - \hat{\delta}_1)^2}{\sum_{t=1}^T K\left(\frac{t/T-r}{h}\right)}\right]^{1/2}, \tag{1.4}$$

where $K : \mathbb{R} \to \mathbb{R}$ denotes the kernel function and h is the bandwidth parameter.

The kernel function K and the bandwidth h_T are assumed to satisfy the following conditions.

Assumption K: *The kernel function $K : \mathbb{R} \to \mathbb{R}$ is symmetric about 0, positive and three times differentiable. Additionally, it satisfies $\int K(x)dx > 0$, $\int |x|K(x)dx < \infty$, $\int |xK'(x)|dx < \infty$, $\lim_{|x|\to\infty} x^2 K(x) \to 0$ and $\lim_{|x|\to\infty} x^4 K'(x) \to 0$, monotonously for sufficiently large $|x|$. The Fourier transformation of K, denoted by ψ exists and satisfies $\int |\lambda\psi(\lambda)|d\lambda < \infty$.*

Assumption H: *The bandwidth parameter satisfies $h_T \to 0$ and $Th_T^4 \to \infty$ as $T \to \infty$.*

The estimator $\hat{g}$ is used to construct the series

$$y_t^* = \sum_{s=1}^{t} \frac{y_s - y_{s-1}}{\hat{g}\left(\frac{s}{T}\right)}. \tag{1.5}$$

It is important to note that the series $(y_t^*)_{t=1,\dots,T}$ is constructed in such a way that its differences are approximately homoskedastic. Then, Beare (2006) suggests to apply the standard Phillips-Perron test to the series $(y_t^*)_{t=1,\dots,T}$. Running the regression of y_t^* on y_{t-1}^* and the deterministic components z_t leads to least squares estimators $\hat{\rho}^*$ and $\hat{\delta}^* = (\hat{\delta}_0^*, \hat{\delta}_1^*)^\top$. Then, it follows that the test statistic is given by

$$Z_{\hat{\rho}^*} = T \cdot (\hat{\rho}^* - 1) - \frac{\hat{\lambda}^*}{T^{-2}\sum_{t=1}^{T} y_{t-1}^{*2}},$$

where $\hat{\lambda}^*$ is defined as

$$\hat{\lambda}^* = \sum_{j=1}^{m} K\left(\frac{j}{m}\right)\left(\frac{1}{T}\sum_{t=1}^{T-j}\left(y_t^* - \hat{\delta}^{*\top} z_t - \hat{\rho}^* y_{t-1}^*\right)\left(y_{t+j}^* - \hat{\delta}^{*\top} z_{t+j} - \hat{\rho}^* y_{t+j-1}^*\right)\right).$$

The asymptotic distribution of the test statistic $Z_{\hat{\rho}^*}$ is stated in the following theorem.

Theorem 1.2:

Suppose Assumptions V, E, K and H hold.

(a) Let $z_t = 1$. As $T \to \infty$,

$$Z_{\rho^*} \Rightarrow \frac{\frac{1}{2}\left[W(1)^2 - 1\right] - W(1)\int_0^1 W(r)dr}{\int_0^1 W(r)^2 dr - \left[\int_0^1 W(r)dr\right]^2}.$$

(b) Let $z_t = (1, t)^\top$. As $T \to \infty$,

$$Z_{\rho^*} \Rightarrow a/\sqrt{b},$$

where

$$\begin{aligned} a &= \int_0^1 W(r)dW(r) - W(1)\int_0^1 W(r)dr \\ &+ 12\left[\int_0^1 rW(r)dr - \frac{1}{2}\int_0^1 W(r)dr\right]\left[\int_0^1 W(r)dr - \frac{1}{2}W(1)\right] \end{aligned}$$

and

$$\begin{aligned} b &= \int_0^1 W(r)^2 dr - 12\left(\int_0^1 rW(r)dr\right)^2 \\ &+ 12\int_0^1 W(r)dr\int_0^1 rW(r)dr - 4\left(\int_0^1 W(r)\right)^2. \end{aligned}$$

The proof of the first part of Theorem 1.2 is given in Beare (2006). The second part of Theorem 1.2 can be established along the lines of Beare (2006).

1.3.2 GLS Detrended Phillips-Perron Test

In this subsection we propose a slight modification to the procedure described in the previous subsection. Let y_t^* be defined as in (1.5). Inspired by Elliott et al. (1996)

we define

$$y_{\bar{c}}^* = \left(y_0^*, y_1^* - \left(1 - \frac{\bar{c}}{T}\right) y_0^*, \ldots, y_T^* - \left(1 - \frac{\bar{c}}{T}\right) y_{T-1}^*\right)^\top$$

and

$$z_{\bar{c}} = \left(z_0, z_1 - \left(1 - \frac{\bar{c}}{T}\right) z_0, \ldots, z_T - \left(1 - \frac{\bar{c}}{T}\right) z_{T-1}\right)^\top.$$

The value of $\bar{c}$ depends on the deterministic components; Elliott et al. (1996) suggest to set $\bar{c} = 7$ when $z_t = 1$ and $\bar{c} = 13.5$ when $z_t = (1, t)^\top$. Next, regress $y_{\bar{c}}^*$ on $z_{\bar{c}}$, denote the corresponding least squares estimator by $\hat{\gamma}$ and define the local GLS residuals $\hat{y}_t^* = y_t^* - \hat{\gamma}^\top z_t$. Then, the test statistic is given by

$$Z_{\rho^*}^{\mathrm{GLS}} = T \cdot (\hat{\rho}_{\mathrm{GLS}}^* - 1) - \frac{\hat{\lambda}_{\mathrm{GLS}}^*}{T^{-2} \sum_{t=1}^T \hat{y}_{t-1}^{*2}}, \tag{1.6}$$

where $\hat{\rho}_{\mathrm{GLS}}^*$ is the least squares estimator of the regression of $\hat{y}_t^*$ on $\hat{y}_{t-1}^*$ and $\hat{\lambda}_{\mathrm{GLS}}^*$ is defined as

$$\hat{\lambda}_{\mathrm{GLS}}^* = \sum_{j=1}^m K\left(\frac{j}{m}\right) \left(\frac{1}{T} \sum_{t=1}^{T-j} \left(\hat{y}_t^* - \hat{\rho}_{\mathrm{GLS}}^* \hat{y}_{t-1}^*\right) \left(\hat{y}_{t+j}^* - \hat{\rho}_{\mathrm{GLS}}^* \hat{y}_{t+j-1}^*\right)\right).$$

The following theorem shows that the asymptotic distribution of the corrected test coincides with the limit distribution that the uncorrected test would have if the innovations were homoskedastic. Therefore, we are able to use the critical values provided in Table 1 in Ng and Perron (2001, p. 1524).

Theorem 1.3:

Suppose Assumptions V, E, K and H hold.

(a) Let $z_t = 1$. As $T \to \infty$,

$$Z_{\rho^*}^{GLS} \Rightarrow \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr},$$

where $W(r)$ is a standard Brownian motion.

(b) Let $z_t = (1,t)^\top$. As $T \to \infty$,

$$Z_{\rho^*}^{GLS} \Rightarrow \frac{\frac{1}{2}(W^{tr}(1)^2 - 1)}{\int_0^1 W^{tr}(r)^2 dr},$$

where $W^{tr}(r) := W(r) - r\left(\nu_{\bar{c}} W(1) + 3(1-\nu_{\bar{c}})\int_0^1 rW(r)dr\right)$ and $\nu_{\bar{c}} := (1+\bar{c})/(1+\bar{c}+\bar{c}^2/3)$.

The proof is based on Lemma 10 in Beare (2006) and follows exactly the same steps as in Theorem 2 of Cavaliere (2004). Hence, it is omitted.

1.4 The M Class of Unit Root Tests

Most of the conventional unit root tests have a substantial size distortion when the errors are serially correlated. For this reason, Stock (1999) proposes a class of unit root tests (the so-called M tests) being robust to autocorrelation. Recently, Cavaliere and Taylor (2008) extend the M tests to the case of potential heteroskedasticity. Their approach can be described as follows.

Similar to the previous subsection, let

$$y_{\bar{c}} = \left(y_0, y_1 - \left(1 - \frac{\bar{c}}{T}\right) y_0, \ldots, y_T - \left(1 - \frac{\bar{c}}{T}\right) y_{T-1}\right)^\top$$

and

$$z_{\bar{c}} = \left(z_0, z_1 - \left(1 - \frac{\bar{c}}{T}\right) z_0, \ldots, z_T - \left(1 - \frac{\bar{c}}{T}\right) z_{T-1}\right)^\top.$$

Denote by $\hat{y}_t$ the local GLS residuals $\hat{y}_t = y_t - \hat{\gamma}^\top z_t$, where $\hat{\gamma}$ is the least squares estimator from the regression of $y_{\bar{c}}$ on $z_{\bar{c}}$. The M unit root statistics are given by

$$\begin{aligned} MZ_\rho &= \frac{T^{-1}\hat{y}_T^2 - s_{AR}^2(k)}{2T^{-2}\sum_{t=1}^{T}\hat{y}_{t-1}^2}, && (1.7) \\ MSB &= \left(T^{-2}\sum_{t=1}^{T}\hat{y}_{t-1}^2/s_{AR}^2(k)\right)^{1/2}, && (1.8) \\ MZ_t &= MZ_\rho \cdot MSB, && (1.9) \end{aligned}$$

where

$$\begin{aligned} s_{AR}^2(k) &= \hat{\sigma}^2/\left(1-\hat{\beta}(1)\right)^2, && (1.10) \\ \hat{\beta}(1) &= \sum_{i=1}^{k}\hat{\beta}_{i,k}. && (1.11) \end{aligned}$$

In (1.11), $\hat{\beta}_{i,k}, i = 1, \ldots, k$, are estimators of the coefficients $\beta_{i,k}$ in the Dickey-Fuller regression

$$\Delta\hat{y}_t = \pi\hat{y}_{t-1} + \sum_{i=1}^{k}\beta_{i,k}\Delta\hat{y}_{t-i} + u_t$$

and in (1.10)

$$\hat{\sigma}^2 = \frac{1}{T-k-1}\sum_{t=1}^{T}\hat{u}_t,$$

with

$$\hat{u}_t = \Delta\hat{y}_t - \hat{\pi}\hat{y}_{t-1} + \sum_{i=1}^{k}\hat{\beta}_{i,k}\Delta\hat{y}_{t-i} \quad (1.12)$$

is an estimator for $\sigma^2 = E(u_t^2)$, respectively.

The main result of Cavaliere and Taylor (2008) is articulated in the following theorem. For simplicity, we focus on the case that $z_t = 1$.

Theorem 1.4: *Suppose Assumptions V, E, K and H hold. Then, as $T \to \infty$*

$$\begin{aligned} MZ_\rho &\Rightarrow \frac{\frac{1}{2}(W_g(1)^2 - 1)}{\int_0^1 W_g(r)^2 dr} =: \xi_1, \\ MSB &\Rightarrow \left(\int_0^1 W_g(r)^2 dr\right)^{1/2} =: \xi_2, \\ MZ_t &\Rightarrow \xi_1 \cdot \xi_2 = \frac{\frac{1}{2}(W_g(1)^2 - 1)}{\left(\int_0^1 W_g(r)^2 dr\right)^{1/2}}. \end{aligned}$$

It is evident from Theorem 1.4 that the asymptotic distribution depends on the time-path of the variance and, hence, general critical values are not available. Cavaliere and Taylor (2008) propose a wild bootstrap algorithm to obtain valid critical values given the time-path of the variance. The algorithm consists of the following steps.
Step 1: Calculate the residuals $\hat{u}_t$ given by (1.12) using the original sample.
Step 2: Generate the bootstrap residuals $u_t^* := \hat{u}_t w_t$, where $(w_t)_{t=1,\ldots,T}$ denotes an independent $N(0,1)$ sequence.
Step 3: The bootstrap sample is obtained by using the model under the null hypothesis

$$y_t^* = y_0^* + \sum_{i=1}^{t} u_i^*, \qquad t = 1, \ldots, T$$

for some initial value y_0^*.
Step 4: Calculate the value of the M test statistics for the bootstrap sample $(y_t^*)_{t=1,\ldots,T}$.
Step 5: Repeat steps 2-4 B times.
Step 6: Reject the null hypothesis if the value of the M test statistics based on the original sample is smaller than an appropriate quantile of the bootstrap distribution generated in the previous step.

Cavaliere and Taylor (2008) establish that under the null hypotheis the described wild bootstrap algorithm leads to tests with asymptotically correct size and, furthermore, the procedure is consistent under the alternative.

1.5 The ML Unit Root Test

Under the assumption that the volatility of the innovations is constant, Elliott, Rothenberg and Stock (1996) derive for standard unit root tests the asymptotic power envelope which is an upper bound for the power function. When the volatility process is non-stationary and deterministic components are excluded, Boswijk (2005) characterizes the power envelope and constructs a class of feasible tests whose power functions are tangent to the power envelope at one point.[1] In particular, Boswijk (2005) observes that the power of the Gaussian ML coefficient test is close to the power envelope. To allow for deterministic components, we employ the local-to-unity GLS approach by Elliott et al. (1996) and define

$$y_{\bar{c}} = \left(y_0, y_1 - \left(1 - \frac{\bar{c}}{T}\right) y_0, \ldots, y_T - \left(1 - \frac{\bar{c}}{T}\right) y_{T-1}\right)^\top$$

and

$$z_{\bar{c}} = \left(z_0, z_1 - \left(1 - \frac{\bar{c}}{T}\right) z_0, \ldots, z_T - \left(1 - \frac{\bar{c}}{T}\right) z_{T-1}\right)^\top,$$

where $\bar{c}$ is defined as in the previous section. Let $\hat{y}_t = y_t - \hat{\gamma}^\top z_t$, where $\hat{\gamma}$ is the least squares estimator from the regression of $y_{\bar{c}}$ on $z_{\bar{c}}$. Similar to Boswijk (2005), we propose the ML test statistic

$$ML = T(\hat{\rho}^*_{\text{GLS}} - 1), \tag{1.13}$$

where $\hat{\rho}^*_{\text{GLS}} = (\sum_{t=1}^T \hat{y}^{*2}_{t-1})^{-1} \sum_{t=1}^T \hat{y}^*_{t-1}\hat{y}^*_t =: J_T^{-1} S_T$[2] is the GLS estimator of the regression of $\hat{y}^*_t$ on $\hat{y}^*_{t-1}$, with $\hat{y}^*_t := \frac{\hat{y}_t}{\hat{g}\left(\frac{t}{T}\right)}$ and $\hat{g}(\cdot)$ is given by (1.4) Note that the GLS

[1] The derivation of the power envelope is based on the Neyman-Pearson lemma, that is the comparison of the log-likelihood of $\rho_T = \left(1 - \frac{c}{T}\right)$, $c > 0$ with the log-likelihood of $\rho = 1$.

[2] Note that S_T is the score function at $\rho = 1$ and J_T is the Fisher Information at $\rho = 1$.

estimator $\hat{\rho}^*_{\text{GLS}}$ coincides with the Gaussian ML estimator. Since the asymptotic distribution depends on the variance path $g(r)$, $r \in [0,1]$, p-values of the ML test are obtained by simulation of S_T and J_T, given $\hat{g}(\cdot)$.

1.6 Finite-sample simulations

1.6.1 Setup

In this section we investigate the finite sample properties of the tests described in the previous sections. The data are generated by the two processes

$$\begin{aligned}
\text{DGP 1:} \quad y_t &= 1 + x_t \\
x_t &= \left(1 - \frac{c}{T}\right) x_{t-1} + g\left(\frac{t}{T}\right) \epsilon_t, \\
\text{DGP 2:} \quad y_t &= 1 + t + x_t \\
x_t &= \left(1 - \frac{c}{T}\right) x_{t-1} + g\left(\frac{t}{T}\right) \epsilon_t,
\end{aligned}$$

where $c \geq 0$. The error process $(\epsilon_t)_{t=1\ldots,T}$ was taken to be a Gaussian white noise process. This allows us to focus on the effect of heteroskedasticity on the tests independent of the effect of serial correlation. We consider ten different choices of g being similar to Beare (2006) and Cavaliere and Taylor (2008). The choices are given in Table 1.1.

Model 1 is homoskedastic. Models 2-4 feature a single variance break. Models 5 and 6 have two variance breaks. Models 7-10 are exponential near-integrated or integarted stochastic volatility models. Note that the latter models are stochastic and, hence, are not covered by our framework. To investigate the size of the tests we set $c = 0$ such that the process $(x_t)_{t=1,\ldots,T}$ has a unit root. The nominal size is set to $\alpha = 5\%$. Under the alternative we set c to 3, 6 and 9. Note that if c differs from zero, $(x_t)_{t=1,\ldots,T}$ is stationary. Local power of the standard Phillips-Perron test and M tests is computed by using the 95% empirical quantile leading to an exact size under the null hypothesis (size-adjusted power). The bootstrap distributions of the M tests of

Table 1.1: Models of heteroskadasticity considered in Tables 1.2 - 1.17

Model	Volatility Function
1	$g(r)^2 = 1$
2	$g(r)^2 = 1\{r < 0.8\} + 9\{r \geq 0.8\}$
3	$g(r)^2 = 1\{r < 0.2\} + \frac{1}{9}\{r \geq 0.2\}$
4	$g(r)^2 = 1\{r < 0.8\} + \frac{1}{9}\{r \geq 0.8\}$
5	$g(r)^2 = 1\{r < 0.2\} + \frac{1}{9}\{0.2 \geq r < 0.8\} + 1\{r \geq 0.8\}$
6	$g(r)^2 = 1\{r < 0.4\} + 9\{0.4 \geq r < 0.6\} + 1\{r \geq 0.6\}$
7	$g(r)^2 = \exp(4J_{-10}(r))$
8	$g(r)^2 = \exp(9J_{-10}(r))$
9	$g(r)^2 = \exp(4B(r))$
10	$g(r)^2 = \exp(9B(r))$

section 1.3 are approximated by $B = 499$ replications. The nonparametric variance estimator (1.4) is computed by using a Gaussian kernel with bandwidth $h = \kappa \cdot T^{-0.2}$. The parameter κ is set to 0.2, 0.4 and 0.6. This enables us to investigate the effect of the bandwidth parameter on size and power of the modified Phillips-Perron test and the ML test. Size and power are computed for the standard Phillips-Perron test (Z_ρ), the modified Phillips-Perron test by Beare ($Z^*_{\rho,\kappa}$) with bandwidth parameter $h = \kappa \cdot T^{-0.2}$, the GLS detrended version of the Phillips-Perron test ($Z^{\text{GLS}}_{\rho^*,\kappa}$) with bandwidth parameter $h = \kappa \cdot T^{-0.2}$, the standard M class of unit root tests (MZ_α, MSB and MZ_t), the bootstrap M tests (MZ^*_α, MSB^* and MZ^*_t) and the ML test (ML^*_κ) with bandwidth parameter $h = \kappa \cdot T^{-0.2}$. Two sample sizes are considered, viz $T = 100$ and $T = 250$. All results are based on 5000 replications.

1.6.2 Results for DGP 1

Size properties

Empirical size under a nominal size of 5% is reported in Table 1.2 for the standard Phillips-Perron test and the Phillips-Perron test by Beare (2006), in Table 1.3 for the GLS detrended variants of the Phillips-Perron test, in Table 1.4 for the M tests

Table 1.2: Size of standard PP test (Z_ρ) and PP test by Beare (Z_ρ^*), DGP 1

	T=100				T=250			
Model	Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.6}$	Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.6}$
1	3.68	3.56	3.52	3.54	5.14	5.06	5.14	5.18
2	7.76	3.60	4.38	5.02	9.08	5.28	5.74	6.18
3	13.70	3.60	4.64	5.90	15.50	5.12	5.72	6.54
4	4.44	3.84	4.02	4.04	5.72	4.74	5.06	5.16
5	10.78	4.22	5.84	7.54	12.46	5.20	6.92	8.00
6	6.70	3.70	5.02	5.74	7.62	5.52	6.06	6.66
7	5.34	4.06	4.30	4.70	6.10	5.16	5.28	5.32
8	10.22	4.16	5.86	7.40	11.76	5.80	7.54	8.38
9	12.44	4.40	5.92	7.58	13.46	5.08	6.18	7.72
10	25.14	6.60	13.58	18.50	26.30	6.86	12.72	18.06

Table 1.3: Size of GLS detrended PP test (Z_ρ^{GLS}) and GLS detrended PP test based on Beare ($Z_{\rho^*}^{\text{GLS}}$), DGP 1

	T=100				T=250			
Model	Z_ρ^{GLS}	$Z^{\text{GLS}}_{\rho^*,0.2}$	$Z^{\text{GLS}}_{\rho^*,0.4}$	$Z^{\text{GLS}}_{\rho^*,0.6}$	Z_ρ^{GLS}	$Z^{\text{GLS}}_{\rho^*,0.2}$	$Z^{\text{GLS}}_{\rho^*,0.4}$	$Z^{\text{GLS}}_{\rho^*,0.6}$
1	7.04	6.92	6.90	6.92	5.76	5.66	5.72	5.72
2	12.64	6.48	7.06	7.26	12.24	6.04	6.34	6.80
3	9.86	6.72	7.40	7.58	8.06	5.76	5.62	6.08
4	7.08	7.02	6.84	6.98	5.60	5.54	5.54	5.72
5	9.58	6.52	7.56	8.30	8.08	5.92	6.28	7.06
6	11.10	6.98	7.64	9.24	9.54	5.78	6.40	7.32
7	8.44	6.76	7.36	7.72	7.44	5.72	6.04	6.24
8	13.32	6.50	8.52	9.98	11.98	6.24	7.44	8.56
9	13.94	7.04	7.94	9.74	11.94	5.70	6.84	7.80
10	22.18	8.12	13.70	17.44	19.26	6.94	10.42	13.86

Table 1.4: Size of standard M tests (MZ_α, MSB, MZ_t) and bootstrap M tests (MZ^*_α, MSB^*, MZ^*_t), DGP 1

	T=100						T=250					
Model	MZ_α	MSB	MZ_t	MZ^*_α	MSB^*	MZ^*_t	MZ_α	MSB	MZ_t	MZ^*_α	MSB^*	MZ^*_t
1	6.44	5.86	5.86	5.12	5.08	5.02	5.70	5.20	5.30	5.18	5.20	5.36
2	11.88	14.48	10.26	5.42	5.30	5.48	12.06	13.22	10.50	5.54	5.68	5.62
3	9.16	8.16	8.82	5.72	5.56	5.68	7.74	7.12	7.32	5.44	5.50	5.40
4	6.40	5.32	6.44	4.90	4.94	5.10	5.40	4.60	5.32	4.76	4.82	4.88
5	8.88	9.82	7.82	5.44	5.34	5.46	7.94	8.48	7.12	5.46	5.44	5.52
6	10.34	9.40	10.00	5.84	5.70	5.92	9.44	8.52	9.04	5.56	5.44	5.64
7	7.80	7.04	7.24	5.56	5.24	5.36	7.26	7.10	6.60	5.54	5.50	5.34
8	12.64	12.96	11.80	6.26	6.16	6.26	11.78	11.64	11.16	5.86	5.98	5.72
9	13.30	13.92	12.22	5.88	5.58	6.02	11.72	11.96	10.74	5.30	5.24	5.46
10	21.56	24.08	19.80	8.14	7.84	8.00	19.08	21.78	17.32	5.94	5.94	5.84

Table 1.5: Size of the ML test, DGP 1

	T=100			T=250		
Model	$ML^{\text{GLS*}}_{0.2}$	$ML^{\text{GLS*}}_{0.4}$	$ML^{\text{GLS*}}_{0.6}$	$ML^{\text{GLS*}}_{0.2}$	$ML^{\text{GLS*}}_{0.4}$	$ML^{\text{GLS*}}_{0.6}$
1	4.34	4.76	4.88	4.72	4.92	4.98
2	4.14	4.88	5.28	5.16	5.46	5.94
3	4.14	4.46	5.00	4.24	4.68	5.08
4	4.54	4.88	4.82	4.44	4.42	4.44
5	3.96	4.70	5.78	4.26	4.90	5.70
6	4.28	5.52	7.06	4.70	5.44	6.40
7	4.38	5.12	5.44	4.80	5.24	5.62
8	4.16	6.20	7.18	4.50	5.88	7.38
9	4.08	5.18	6.42	4.56	5.62	6.62
10	3.82	6.04	9.78	3.56	5.18	8.38

and in Table 1.5 for the ML test, respectively. The most interesting feature is that all uncorrected versions of the tests have substantial size distortion for models 2, 3, 5, 6, 8, 9 and, in particular, model 10. Depending on the test considered the latter model yields a size between 17.32 (MZ_t test when $T = 250$) and 26.30 (standard Phillips-Perron test when $T = 250$).

Let us initially turn to the results of the Phillips-Perron test by Beare (2006). Table 1.2 shows that the selection of the bandwidth parameter has a significant effect on the empirical size. Whereas both $\kappa = 0.2$ and $\kappa = 0.4$ lead to a size much closer to the nominal size, the choice $\kappa = 0.6$ suffers from higher over-rejection.

The results for the GLS detrended versions of the Phillips-Perron test are essentially comparable with those in Table 1.2. However, the results shown in Table 1.3 indicate that the $Z_{\rho^*}^{\mathrm{GLS}}$-test has generally a slightly greater size distortion than the Z_{ρ}^{*}-test.

The size properties of the standard M tests and the bootstrap M test are given in Table 1.4. We note that the bootstrap M tests perform very well; they feature sizes between 4% and 6% in eight of ten cases when $T = 100$ and in all cases when $T = 250$. It is worth noting that the differences between the MZ_{α}-test, the MSB-test and MZ_t-test are negligible.

Finally, consider the size properties of the ML unit root test presented in Table 1.5. The empirical size of the ML test depends on the choice of the bandwidth parameter similar to the Phillips-Perron test. When $\kappa = 0.2$, empirical size is very close to the nominal size; when $T = 250$ size peaks at 5.16% for model 2 and reaches a low at 3.56% for model 10; results for $T = 100$ are similar. If $\kappa = 0.4$, sizes are between 4% and 6% in all cases when $T = 250$. Setting $\kappa = 0.6$, the size distortion is more severe, in particular for models 6, 8, 9 and 10. In these cases the size distortion is lowered by choosing a smaller bandwidth, e.g. $\kappa = 0.2$.

Power properties

Local power results are presented in Table 1.6 for the standard Phillips-Perron test and the Phillips-Perron test by Beare (2006), in Table 1.7 for the GLS detrended

Table 1.6: Power of standard PP test (Z_ρ) and PP test by Beare (Z_ρ^*), DGP 1

Model		T=100				T=250			
		Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.6}$	Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.6}$
1	c=3	13.64	13.62	14.02	14.08	11.12	11.18	11.26	11.18
	c=6	26.66	26.82	26.90	27.08	22.00	21.82	21.98	21.74
	c=9	45.86	45.56	46.04	46.32	37.90	37.08	37.74	37.94
2	c=3	11.90	12.66	11.92	12.30	12.06	10.84	11.26	11.76
	c=6	21.30	25.10	23.98	23.66	21.26	20.90	21.60	22.48
	c=9	33.88	42.18	40.74	40.44	33.76	35.20	36.00	37.18
3	c=3	9.72	8.64	8.58	8.70	8.18	6.34	6.44	7.20
	c=6	17.42	15.56	15.94	16.42	14.38	11.78	11.82	13.00
	c=9	28.82	27.90	29.04	29.70	23.80	22.02	22.50	23.80
4	c=3	13.44	12.56	13.10	13.54	11.58	11.18	11.42	11.66
	c=6	26.18	24.04	25.34	26.04	22.00	21.10	22.50	22.76
	c=9	42.64	40.94	42.38	42.80	38.06	35.76	38.06	38.56
5	c=3	9.80	8.26	8.46	9.02	8.72	6.76	7.26	7.56
	c=6	16.82	14.42	15.30	15.86	15.54	12.14	13.14	13.62
	c=9	27.66	25.54	26.48	27.28	25.00	21.26	22.58	23.28
6	c=3	12.46	13.52	13.72	13.76	11.86	11.08	12.84	12.44
	c=6	23.46	24.16	26.72	26.00	21.88	20.18	24.00	24.02
	c=9	37.22	38.48	43.50	41.98	35.12	31.08	38.44	38.30
7	c=3	12.70	11.84	12.50	12.40	12.28	10.90	11.72	12.16
	c=6	23.32	22.02	23.02	22.98	22.34	20.72	22.06	22.82
	c=9	39.44	37.44	39.16	39.24	37.98	33.96	36.84	37.94
8	c=3	10.32	10.92	11.20	10.68	9.92	8.68	11.14	11.46
	c=6	17.04	18.88	19.86	18.76	17.32	14.52	18.60	19.38
	c=9	26.92	29.32	32.54	30.98	26.80	22.76	30.66	31.42
9	c=3	8.02	9.78	9.38	9.48	6.80	9.28	9.16	8.34
	c=6	13.76	17.96	17.92	17.24	12.04	16.90	17.56	16.16
	c=9	22.90	30.80	31.14	29.30	19.92	28.20	30.22	28.30
10	c=3	5.58	5.92	5.34	4.92	4.50	5.94	4.24	3.82
	c=6	7.96	9.70	9.22	7.46	6.66	9.66	7.78	6.88
	c=9	10.56	16.48	14.76	12.24	9.58	15.24	13.40	11.62

Table 1.7: Power of GLS detrended PP test (Z_ρ^{GLS}) and GLS detrended PP test based on Beare ($Z_{\rho^*}^{\text{GLS}}$), DGP 1

		T=100				T=250			
Model		Z_ρ^{GLS}	$Z_{\rho^*,0.2}^{\text{GLS}}$	$Z_{\rho^*,0.4}^{\text{GLS}}$	$Z_{\rho^*,0.6}^{\text{GLS}}$	Z_ρ^{GLS}	$Z_{\rho^*,0.2}^{\text{GLS}}$	$Z_{\rho^*,0.4}^{\text{GLS}}$	$Z_{\rho^*,0.6}^{\text{GLS}}$
1	c=3	15.96	15.48	15.64	15.84	16.54	16.46	16.60	16.50
	c=6	36.96	34.66	35.76	36.20	38.42	37.06	38.28	38.20
	c=9	60.66	55.96	58.62	59.80	62.94	61.14	62.28	62.58
2	c=3	14.88	15.58	14.46	15.50	15.50	14.92	14.24	15.06
	c=6	30.10	34.06	32.12	33.64	31.14	32.90	31.96	32.72
	c=9	47.28	54.32	52.70	54.90	47.70	54.22	54.26	53.84
3	c=3	14.22	13.18	12.86	14.08	16.04	13.94	16.18	16.62
	c=6	23.70	23.94	26.68	27.90	30.96	25.24	32.10	33.66
	c=9	36.10	40.22	45.10	47.86	46.90	40.62	51.94	54.98
4	c=3	16.16	17.08	16.80	16.88	17.66	18.42	18.90	18.60
	c=6	36.28	35.92	36.74	37.26	39.94	39.36	42.68	42.20
	c=9	59.50	56.40	60.00	60.74	64.10	61.48	65.78	65.66
5	c=3	13.84	13.84	14.92	14.16	15.10	13.64	15.70	15.82
	c=6	26.38	24.72	28.06	27.66	29.70	24.00	29.70	31.24
	c=9	41.02	39.96	45.64	44.36	47.28	38.16	47.18	49.82
6	c=3	14.84	14.94	16.22	15.44	15.52	15.14	16.06	15.86
	c=6	30.68	28.82	34.10	32.74	32.70	29.48	33.78	35.12
	c=9	48.80	45.32	53.60	52.34	53.18	45.42	53.10	55.84
7	c=3	15.96	15.00	15.88	15.80	16.00	16.46	17.86	18.18
	c=6	33.60	31.08	34.30	34.10	35.18	34.60	38.10	38.08
	c=9	54.62	49.20	55.32	56.02	56.60	53.74	60.50	60.96
8	c=3	12.44	13.80	14.88	14.28	14.02	12.86	14.94	15.66
	c=6	23.88	24.92	29.24	28.02	26.76	23.16	29.58	31.60
	c=9	37.80	38.74	47.40	44.62	42.16	35.78	47.28	49.44
9	c=3	10.78	12.44	13.40	13.12	11.40	14.08	14.92	14.06
	c=6	19.26	24.58	28.54	26.32	22.60	26.34	30.24	29.94
	c=9	31.00	39.66	44.16	43.20	36.10	41.46	48.50	47.96
10	c=3	7.90	7.90	8.58	7.86	8.54	7.96	8.40	8.52
	c=6	11.78	13.64	15.24	13.34	13.30	12.86	15.80	15.44
	c=9	16.70	22.52	25.24	20.24	19.14	19.80	25.44	24.40

Table 1.8: Power of standard M tests (MZ_α, MSB, MZ_t) and bootstrap M tests (MZ_α^*, MSB^*, MZ_t^*), DGP 1

		T=100						T=250					
Model		MZ_α	MSB	MZ_t	MZ_α^*	MSB^*	MZ_t^*	MZ_α	MSB	MZ_t	MZ_α^*	MSB^*	MZ_t^*
1	c=3	15.78	14.92	16.10	16.14	15.12	16.32	16.62	16.16	16.54	17.50	16.48	18.12
	c=6	36.34	33.22	36.68	36.38	33.54	37.06	38.46	36.60	37.72	39.22	37.42	39.74
	c=9	60.12	56.52	60.44	59.68	56.68	60.66	62.94	60.60	62.52	64.30	61.36	64.70
2	c=3	14.84	13.78	14.50	15.68	15.04	15.80	15.48	14.26	15.56	16.94	16.48	16.94
	c=6	30.16	28.36	28.88	31.60	30.46	31.20	31.10	30.52	30.70	32.90	33.32	32.30
	c=9	47.30	45.10	44.90	48.70	47.50	47.26	47.80	47.90	46.60	50.16	51.96	48.80
3	c=3	13.92	13.54	14.56	15.26	14.80	15.78	15.92	15.12	16.20	16.48	15.56	16.94
	c=6	22.84	22.00	23.70	26.06	24.20	26.42	30.72	28.68	31.80	31.44	29.90	32.46
	c=9	34.72	32.54	36.34	37.82	36.08	39.22	46.46	43.58	48.18	47.06	44.88	48.44
4	c=3	16.08	15.10	17.08	16.50	15.48	17.28	17.60	16.24	18.02	17.32	16.24	17.82
	c=6	35.88	33.34	37.82	35.64	33.80	37.44	39.94	37.48	41.38	39.26	36.80	41.00
	c=9	58.60	54.92	61.74	58.40	55.56	60.70	63.96	60.42	65.76	62.60	59.70	64.50
5	c=3	13.78	12.86	13.88	15.00	13.86	15.52	14.90	14.98	14.94	16.66	16.16	16.98
	c=6	25.82	23.98	25.92	27.68	25.38	28.20	29.42	28.80	29.14	31.94	30.80	32.20
	c=9	39.94	37.20	39.92	41.68	39.16	41.58	46.88	45.36	45.52	49.50	48.06	48.44
6	c=3	14.48	14.14	15.22	17.04	16.46	17.18	15.50	15.26	15.34	16.58	15.90	17.04
	c=6	30.44	28.80	31.90	33.70	31.84	35.02	32.62	30.80	32.88	34.58	32.52	35.48
	c=9	48.26	45.44	50.60	52.50	49.82	54.04	53.04	50.88	53.22	53.88	51.72	55.10
7	c=3	16.02	15.84	16.34	16.72	15.72	17.38	15.96	15.38	16.06	17.62	16.88	17.98
	c=6	33.86	33.36	34.54	35.00	33.00	35.58	35.10	33.16	36.44	37.76	35.80	37.88
	c=9	54.74	53.62	55.70	56.24	52.98	57.12	56.46	53.68	56.94	59.50	56.70	59.94
8	c=3	12.56	11.92	13.26	17.46	16.24	18.08	14.00	13.44	14.32	17.00	16.32	17.32
	c=6	23.90	22.44	25.08	31.94	30.52	32.88	26.74	24.56	28.12	32.42	31.24	33.20
	c=9	37.86	34.76	40.00	48.12	45.88	48.96	42.16	39.64	43.78	50.04	47.78	50.80
9	c=3	10.78	9.84	11.10	15.38	14.24	15.44	11.44	10.46	11.90	14.58	13.96	14.84
	c=6	19.26	17.06	19.84	27.38	26.00	27.66	22.58	20.12	24.16	28.92	28.00	29.48
	c=9	30.84	27.10	31.96	41.80	40.04	42.74	36.02	32.02	37.80	44.68	43.38	45.36
10	c=3	7.76	7.00	8.10	14.76	14.12	14.74	8.52	8.22	8.26	12.06	12.02	12.34
	c=6	11.44	9.82	12.52	23.04	21.82	23.22	13.28	11.98	13.04	19.94	19.42	20.48
	c=9	16.40	12.60	17.78	31.56	30.14	31.42	19.12	16.58	18.96	29.44	28.80	30.12

Table 1.9: Power of the ML test, DGP 1

Model		T=100			T=250		
		$ML_{0.2}^{\mathrm{GLS}*}$	$ML_{0.4}^{\mathrm{GLS}*}$	$ML_{0.6}^{\mathrm{GLS}*}$	$ML_{0.2}^{\mathrm{GLS}*}$	$ML_{0.4}^{\mathrm{GLS}*}$	$ML_{0.6}^{\mathrm{GLS}*}$
1	c=3	14.12	14.94	14.96	15.84	16.48	16.68
	c=6	32.50	33.98	34.50	37.20	37.90	38.20
	c=9	55.22	57.00	57.84	61.20	62.54	62.44
2	c=3	12.70	14.20	15.52	14.74	15.64	16.80
	c=6	28.94	31.50	33.92	34.40	35.76	37.06
	c=9	49.36	52.72	54.32	56.44	57.16	58.62
3	c=3	12.92	13.94	14.44	16.46	16.60	17.00
	c=6	22.48	23.66	24.86	36.36	35.28	35.38
	c=9	33.10	35.22	37.34	53.82	52.38	52.12
4	c=3	15.78	16.12	16.46	17.88	17.56	17.24
	c=6	36.28	36.38	36.26	43.74	41.62	40.90
	c=9	59.64	59.90	59.84	69.22	65.74	64.82
5	c=3	12.32	14.32	16.24	15.40	16.18	17.92
	c=6	24.12	26.98	30.84	34.62	34.64	36.06
	c=9	36.88	42.08	47.62	52.72	53.50	55.56
6	c=3	15.74	18.20	21.38	16.14	17.62	19.66
	c=6	36.98	38.80	42.40	40.28	40.54	43.56
	c=9	61.36	62.96	64.30	68.42	66.86	67.82
7	c=3	14.54	16.06	17.22	16.74	17.80	18.62
	c=6	33.22	35.92	36.62	38.48	39.66	40.60
	c=9	55.80	57.98	58.92	63.10	62.86	63.32
8	c=3	15.52	19.02	21.68	18.20	19.88	21.94
	c=6	37.64	38.16	40.76	47.16	43.06	43.96
	c=9	59.96	59.56	60.96	72.00	65.02	64.54
9	c=3	13.64	15.30	17.70	16.74	17.04	18.68
	c=6	26.60	29.40	32.38	34.68	34.50	36.64
	c=9	41.04	44.26	48.08	53.16	52.10	53.86
10	c=3	18.48	16.76	20.66	28.60	22.54	21.22
	c=6	27.66	26.24	32.22	39.94	32.52	34.56
	c=9	33.96	35.76	43.96	45.90	42.66	47.54

versions of the Phillips-Perron test, in Table 1.8 for the standard M tests and the bootstrap M tests and in Table 1.9 for the ML test, respectively. It is important to note that in general the corrected versions of the tests display a similar power as the standard tests when the variance is constant. Hence, appealingly, the loss in power is negligible when the corrected versions of the tests are applied to homoskedastic data. Under heteroskedasticity the standard tests exhibit a considerably lower power than under homoskedasticity. In particular, the power differs clearly in those cases where the tests suffer from a high size distortion (i.e. model 9 and 10).

Focusing on $T = 250$ we observe that the Z_ρ^*-test exhibits a lower power than the remaining tests. For most volatility models, the $Z_{\rho^*}^{\text{GLS}}$-test, the bootstrap M tests and the $ML^{\text{GLS}\,*}$-test are twice as powerful as the Z_ρ^*-test. Therefore, the results indicate that the local-to-unity GLS approach increases the power substantially. Comparing the $Z_{\rho^*}^{\text{GLS}}$-test with the bootstrap M tests, we observe that the results are ambiguous. Whereas the $Z_{\rho^*}^{\text{GLS}}$-test appears to exhibit a slightly greater power than the bootstrap M tests for models 3 and 4 (both with $\kappa = 0.4$), the pattern is vice versa for models 8 and 9. For model 10, we observe that the bootstrap M tests beats the $Z_{\rho^*}^{\text{GLS}}$-test by ten percentage points. It appears that the power of the ML test is more robust to the choice of the bandwidth parameter than the $Z_{\rho^*}^{\text{GLS}}$-test. All three tests have similar power when the data are homoskedastic, but the ML test is superior to the $Z_{\rho^*}^{\text{GLS}}$-test and the bootstrap M tests for the remaining models. In particular, the differences in power are greater when the volatility function is stochastic (model 7-10) than for a deterministic function (model 1-6). These results are consistent with the findings in Boswijk (2005).

1.6.3 Results for DGP 2

Size properties

Size calculations for DGP 2 are reported in Table 1.10 for the standard Phillips-Perron test and the Phillips-Perron test by Beare (2006), in Table 1.11 for the detrended GLS variants of the Phillips-Perron test, in Table 1.12 for the standard M tests and the

Table 1.10: Size of standard PP test (Z_ρ) and PP test by Beare (Z_ρ^*), DGP 2

	T=100				T=250			
Model	Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.6}$	Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.6}$
1	3.54	3.18	3.34	3.32	4.94	4.34	4.62	4.68
2	7.38	2.96	3.72	4.20	9.58	3.98	4.64	5.40
3	13.52	2.68	3.52	4.36	15.36	3.88	4.62	5.36
4	4.50	2.70	3.42	3.90	5.72	3.74	4.58	4.98
5	9.46	3.14	4.18	5.98	11.06	3.92	5.28	6.70
6	7.16	2.74	4.50	5.62	8.44	3.88	5.20	6.50
7	4.72	2.56	3.46	3.92	6.70	4.62	5.36	5.64
8	9.86	2.64	5.24	6.88	11.58	3.94	5.92	7.68
9	12.14	2.06	3.42	5.04	13.46	3.16	4.06	5.14
10	24.92	1.50	4.82	8.88	25.82	1.56	3.96	7.08

Table 1.11: Size of GLS detrended PP test (Z_ρ^{GLS}) and GLS detrended PP test based on Beare ($Z_{\rho^*}^{\text{GLS}}$), DGP 2

	T=100				T=250			
Model	Z_ρ^{GLS}	$Z^{\text{GLS}}_{\rho^*,0.2}$	$Z^{\text{GLS}}_{\rho^*,0.4}$	$Z^{\text{GLS}}_{\rho^*,0.6}$	Z_ρ^{GLS}	$Z^{\text{GLS}}_{\rho^*,0.2}$	$Z^{\text{GLS}}_{\rho^*,0.4}$	$Z^{\text{GLS}}_{\rho^*,0.6}$
1	5.80	5.10	5.44	5.44	4.68	4.62	4.60	4.66
2	11.00	4.28	5.24	5.90	10.54	3.86	4.42	4.76
3	13.02	4.26	5.20	5.90	8.66	3.92	4.16	4.48
4	6.30	4.22	5.16	5.68	5.40	3.80	4.42	4.76
5	10.18	4.10	5.58	6.98	7.74	3.72	4.32	5.34
6	10.24	4.54	5.96	8.50	9.90	4.38	5.40	6.50
7	7.42	4.92	5.94	6.66	6.44	4.74	5.46	5.74
8	12.72	4.46	7.04	9.10	12.30	4.02	6.16	7.82
9	13.48	3.56	4.96	6.70	10.90	2.94	3.88	5.06
10	23.78	2.44	5.58	9.96	19.66	1.52	3.88	5.96

Table 1.12: Size of standard M tests (MZ_α, MSB, MZ_t) and bootstrap M tests (MZ_α^*, MSB^*, MZ_t^*), DGP 2

	T=100						T=250					
Model	MZ_α	MSB	MZ_t	MZ_α^*	MSB^*	MZ_t^*	MZ_α	MSB	MZ_t	MZ_α^*	MSB^*	MZ_t^*
1	3.00	3.26	2.98	4.84	4.84	4.76	3.70	4.00	3.70	5.14	5.24	5.02
2	7.52	8.90	7.26	6.04	5.60	6.10	9.60	10.64	8.94	5.84	5.96	5.74
3	5.86	6.00	5.96	5.96	6.02	5.82	6.44	6.52	6.62	5.30	5.48	5.34
4	3.36	3.08	3.50	4.98	5.00	4.92	4.58	4.40	4.64	5.08	5.10	5.12
5	5.38	6.00	5.10	5.72	5.56	5.66	5.82	6.70	5.60	5.04	5.16	5.08
6	7.46	7.62	7.38	6.62	6.38	6.50	8.86	8.88	8.76	5.88	5.94	5.94
7	4.40	4.66	4.28	5.48	5.34	5.54	5.62	5.84	5.66	5.48	5.50	5.64
8	9.92	10.36	9.74	7.10	7.06	7.20	11.10	11.78	10.94	6.02	6.00	6.02
9	8.92	9.54	8.46	6.58	6.54	6.60	9.28	9.82	9.14	5.46	5.60	5.40
10	17.36	18.90	16.44	8.92	8.88	8.88	17.52	18.96	17.02	6.30	6.50	6.16

Table 1.13: Size of the ML test, DGP 2

	T=100			T=250		
Model	$ML_{0.2}^{\text{GLS*}}$	$ML_{0.4}^{\text{GLS*}}$	$ML_{0.6}^{\text{GLS*}}$	$ML_{0.2}^{\text{GLS*}}$	$ML_{0.4}^{\text{GLS*}}$	$ML_{0.6}^{\text{GLS*}}$
1	5.22	5.28	5.38	4.84	4.82	4.86
2	5.36	6.22	6.92	5.52	5.86	6.28
3	4.98	5.78	6.18	4.66	4.78	5.20
4	4.72	5.24	5.68	4.54	4.86	5.08
5	4.98	5.72	6.94	4.78	5.32	5.66
6	4.98	6.46	8.40	5.42	6.26	7.50
7	5.12	6.36	6.72	5.46	6.04	6.34
8	5.16	7.94	9.36	4.74	7.24	8.96
9	5.04	6.08	7.40	4.70	5.54	6.40
10	4.58	7.88	12.18	3.84	5.98	8.96

bootstrap M tests and in Table 1.13 for the ML test, respectively. All uncorrected tests exhibit a substantial size distortion for models 2, 3, 5, 6, 8, 9 and 10. In particular for model 10, the standard Phillips-Perron test yields an empirical size of about 26% when $T = 250$.

The most striking aspect is that the corrected Phillips-Perron tests achieve an empirical size being close to the nominal size by choosing a greater bandwidth parameter than for DGP 1. The results propose to set $\kappa = 0.4$ for both the Z_ρ^* test ($\kappa = 0.2$ for DGP 1) and the $Z_{\rho^*}^{\text{GLS}}$-test ($\kappa = 0.2$ for DGP 1). When $\kappa = 0.2$, the empirical size of the ML test is very close to the nominal size; when $T = 250$ sizes are between 4% and 6% in nine of ten cases; results for $T = 100$ are similar. If $\kappa = 0.4$, the test tends to a slight over-rejection; when $T = 250$, the empirical size is about 6% in five of ten cases and even greater when $T = 100$. Setting $\kappa = 0.6$, the size distortion is even worse, in particular for models 2, 6 and 8-10. In these cases the size distortion is lowered by choosing a smaller bandwidth.

Power properties

Simulated power of the tests is quoted in Table 1.14 for the standard Phillips-Perron test and the Phillips-Perron test by Beare (2006), in Table 1.15 for the detrended GLS versions of the Phillips-Perron test, in Table 1.16 for the standard M tests and the bootstrap M tests and in Table 1.17 for the ML test, respectively.

The results show that the power of all tests is lower for DGP 1 than for DGP 2 which is consistent with the findings of Elliott et al (1996). As for DGP 1, we observe that the Z_ρ^*-test is inferior to the Z_ρ^{GLS}-test, the bootstrap M tests and the ML test. Whereas the bootstrap M tests appear to be slightly more powerful than the $Z_{\rho^*}^{\text{GLS}}$-test for DGP 1, Tables 1.15 and 1.16 indicate that the difference in power between the two tests is negligible for DGP 2. For all models the ML test is superior to the $Z_{\rho^*}^{\text{GLS}}$-test and the bootstrap tests.

Table 1.14: Power of standard PP test (Z_ρ) and PP test by Beare (Z^*_ρ), DGP 2

Model		T=100				T=250			
		Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.4}$	Z_ρ	$Z^*_{\rho,0.2}$	$Z^*_{\rho,0.4}$	$Z^*_{\rho,0.6}$
1	c=3	7.28	7.30	7.40	7.48	7.04	7.28	7.34	7.40
	c=6	13.32	13.02	13.18	13.24	11.72	12.06	12.00	12.12
	c=9	23.06	22.22	22.80	23.10	19.24	19.42	19.74	20.06
2	c=3	7.06	7.82	7.46	6.94	7.64	7.82	7.66	7.76
	c=6	11.84	13.68	13.06	12.92	12.14	12.72	12.62	12.62
	c=9	19.92	23.18	22.40	21.82	19.46	20.66	20.36	20.66
3	c=3	6.86	7.04	7.24	7.34	6.22	6.90	7.04	6.96
	c=6	11.34	12.24	13.30	13.08	9.82	10.92	10.92	11.46
	c=9	18.26	20.52	21.94	21.28	15.06	17.40	19.06	18.76
4	c=3	6.30	6.48	6.38	6.36	5.80	5.86	5.70	5.60
	c=6	11.02	11.62	11.22	11.40	10.06	9.98	9.60	9.60
	c=9	18.58	20.40	19.52	19.28	16.78	17.12	16.58	16.66
5	c=3	7.48	7.16	7.40	7.32	7.28	6.78	6.96	7.14
	c=6	12.96	12.70	13.10	12.94	11.64	11.06	11.58	11.12
	c=9	20.98	21.40	21.24	21.40	17.98	17.84	18.98	18.62
6	c=3	6.78	6.80	6.66	6.84	6.76	7.02	7.10	6.68
	c=6	11.02	11.84	11.80	11.32	10.58	11.44	11.06	11.00
	c=9	17.98	19.94	20.32	19.22	16.36	18.70	19.02	18.10
7	c=3	7.34	7.06	6.92	6.82	7.02	6.96	7.20	7.02
	c=6	13.04	12.88	12.46	12.44	11.64	10.94	11.44	11.74
	c=9	21.42	21.54	21.26	21.32	18.76	17.32	18.30	18.76
8	c=3	6.28	7.14	7.14	6.88	6.48	6.90	7.16	6.62
	c=6	9.62	11.88	11.46	11.10	9.44	11.14	11.88	10.62
	c=9	14.74	19.94	18.62	17.54	14.96	17.60	19.14	16.94
9	c=3	6.26	7.66	7.40	7.68	6.74	7.30	7.96	8.06
	c=6	9.04	12.78	13.00	12.34	10.18	11.38	12.62	12.56
	c=9	13.42	21.02	21.46	20.62	14.48	18.82	19.78	19.98
10	c=3	5.48	6.42	6.58	6.22	5.58	6.62	7.10	7.56
	c=6	6.54	10.10	10.08	9.20	6.42	10.10	11.44	12.04
	c=9	8.12	15.58	15.24	13.94	7.92	15.10	17.66	17.18

Table 1.15: Power of GLS detrended PP test (Z_{ρ}^{GLS}) and GLS detrended PP test based on Beare ($Z_{\rho^*}^{\text{GLS}}$), DGP 2

		T=100				T=250			
Model		Z_{ρ}^{GLS}	$Z_{\rho^*,0.2}^{\text{GLS}}$	$Z_{\rho^*,0.4}^{\text{GLS}}$	$Z_{\rho^*,0.6}^{\text{GLS}}$	Z_{ρ}^{GLS}	$Z_{\rho^*,0.2}^{\text{GLS}}$	$Z_{\rho^*,0.4}^{\text{GLS}}$	$Z_{\rho^*,0.6}^{\text{GLS}}$
1	c=3	7.38	6.80	6.80	7.08	7.70	7.84	7.94	8.08
	c=6	14.12	13.54	13.74	13.96	14.56	14.54	14.46	14.90
	c=9	25.58	24.60	24.88	25.30	26.26	26.04	26.64	26.78
2	c=3	7.18	8.10	7.74	7.88	7.78	8.42	8.40	8.54
	c=6	13.08	15.14	13.82	14.10	14.04	15.58	15.30	15.36
	c=9	22.94	26.54	24.74	24.96	23.18	26.94	26.44	25.70
3	c=3	7.10	7.16	7.04	7.20	6.88	7.12	7.16	7.16
	c=6	12.52	13.32	13.04	13.50	12.60	13.40	13.96	14.02
	c=9	20.68	24.60	24.06	24.92	21.34	24.78	25.88	25.86
4	c=3	6.62	6.28	6.24	6.44	6.48	6.42	6.26	6.38
	c=6	12.12	11.96	11.66	11.76	12.28	12.78	12.32	12.38
	c=9	22.96	23.70	22.92	22.68	22.46	23.88	23.02	22.66
5	c=3	7.42	7.66	7.48	7.58	8.04	7.76	7.50	7.54
	c=6	12.94	14.62	14.38	13.78	14.34	15.04	14.60	14.14
	c=9	22.26	24.84	24.76	24.16	24.26	26.92	26.04	24.68
6	c=3	6.98	7.22	7.70	7.24	7.20	6.90	6.72	6.80
	c=6	12.04	13.22	13.64	13.26	11.76	12.68	13.06	12.62
	c=9	19.76	23.30	24.04	22.06	19.68	22.26	22.72	21.50
7	c=3	7.24	7.16	7.22	7.00	7.22	6.90	7.04	7.32
	c=6	13.08	13.50	13.70	13.16	13.28	13.14	13.10	13.54
	c=9	23.20	24.20	24.64	24.20	23.22	23.60	23.50	23.78
8	c=3	6.20	6.94	6.90	7.16	6.48	7.36	6.96	7.10
	c=6	9.92	13.10	12.76	11.84	10.66	12.74	12.64	12.58
	c=9	15.60	22.34	20.78	19.78	16.88	21.74	21.98	20.86
9	c=3	6.38	7.70	7.72	7.44	7.44	8.18	7.58	7.72
	c=6	10.38	13.94	14.14	13.26	12.78	14.12	14.06	13.84
	c=9	16.04	23.74	24.14	23.04	19.18	24.48	24.28	23.88
10	c=3	5.96	6.82	6.20	6.14	6.42	7.22	7.36	7.74
	c=6	7.64	10.90	10.74	9.68	9.14	11.02	12.56	12.54
	c=9	9.64	17.58	16.68	15.02	11.98	17.46	20.00	19.58

Table 1.16: Power of standard M tests (MZ_α, MSB, MZ_t) and bootstrap M tests (MZ_α^*, MSB^*, MZ_t^*), DGP 2

		T=100						T=250					
Model		MZ_α	MSB	MZ_t	MZ_α^*	MSB^*	MZ_t^*	MZ_α	MSB	MZ_t	MZ_α^*	MSB^*	MZ_t^*
1	c=3	7.30	7.32	7.52	7.36	7.42	7.32	7.54	7.64	7.66	7.76	8.02	7.70
	c=6	14.22	13.92	14.54	14.04	13.74	13.92	14.40	13.88	14.56	14.68	14.24	14.62
	c=9	25.56	25.42	26.26	25.20	24.74	25.16	26.18	25.22	26.54	26.62	25.84	26.86
2	c=3	7.16	7.30	7.36	8.86	8.68	8.90	7.72	7.68	7.84	9.00	9.22	8.98
	c=6	12.92	13.08	13.00	15.32	14.68	15.24	14.00	14.08	14.22	15.76	15.80	15.78
	c=9	22.70	22.88	22.72	25.98	25.72	25.62	22.96	22.58	23.68	25.86	26.02	25.66
3	c=3	6.96	6.98	7.28	8.42	8.22	8.44	6.82	6.84	7.12	7.64	7.60	7.60
	c=6	12.16	11.98	12.68	13.92	13.50	14.26	12.54	12.72	12.82	13.34	13.48	13.42
	c=9	19.92	19.48	20.50	22.44	21.94	22.60	21.24	21.00	21.96	22.36	22.02	22.50
4	c=3	6.68	6.76	6.34	6.44	6.66	6.50	6.56	6.46	6.68	6.66	6.70	6.70
	c=6	12.72	12.62	12.32	12.34	12.04	12.22	12.36	12.10	12.62	12.64	12.66	12.62
	c=9	23.42	22.84	23.24	22.34	21.92	22.64	22.60	21.96	23.10	22.94	22.40	23.20
5	c=3	7.56	7.60	7.32	8.54	8.24	8.40	8.16	8.04	8.12	8.46	8.28	8.38
	c=6	13.66	13.38	13.62	15.02	14.50	15.10	14.32	14.00	14.34	14.18	14.12	14.28
	c=9	22.68	22.40	22.50	24.64	23.82	24.56	24.22	24.12	24.76	24.60	24.12	24.72
6	c=3	7.14	7.10	6.90	8.88	8.72	8.70	7.18	7.20	7.14	8.26	8.12	8.30
	c=6	12.22	12.04	12.30	14.64	14.24	14.76	11.76	11.94	11.84	13.56	13.64	13.56
	c=9	19.80	19.62	20.32	22.78	22.46	22.92	19.70	19.70	19.78	22.56	22.34	22.66
7	c=3	7.20	6.80	7.04	7.68	7.38	7.58	7.38	7.14	7.12	8.24	8.24	8.20
	c=6	13.00	12.24	13.34	14.12	13.42	14.26	13.28	12.66	12.96	14.42	14.52	14.58
	c=9	23.56	22.56	24.00	24.56	24.16	25.00	23.34	22.12	22.88	25.48	25.44	25.38
8	c=3	6.26	6.22	6.30	9.40	9.42	9.38	6.48	6.66	6.70	8.34	8.50	8.42
	c=6	9.90	10.00	10.00	15.56	15.02	15.62	10.60	11.10	11.06	13.72	13.78	13.82
	c=9	15.74	15.30	15.64	23.72	23.32	24.06	16.80	17.22	17.44	21.52	21.36	21.68
9	c=3	6.68	6.58	6.86	9.08	9.00	9.18	7.30	7.32	7.58	7.66	7.72	7.60
	c=6	10.60	10.54	10.62	14.72	14.48	14.96	12.82	12.56	12.68	13.50	13.40	13.48
	c=9	16.60	16.22	17.02	23.46	22.76	23.60	19.42	18.88	19.50	21.86	21.90	21.88
10	c=3	6.28	6.02	6.12	10.94	11.04	10.86	6.58	6.52	6.62	8.66	8.60	8.54
	c=6	8.12	7.76	8.04	14.88	14.90	14.86	9.40	9.00	9.48	12.78	12.78	12.88
	c=9	11.02	10.24	11.08	20.12	20.08	20.06	12.20	11.76	12.64	18.24	18.10	18.14

Table 1.17: Power of the ML test, DGP 2

Model		T=100			T=250		
		$ML_{0.2}^{\mathrm{GLS}*}$	$ML_{0.4}^{\mathrm{GLS}*}$	$ML_{0.6}^{\mathrm{GLS}*}$	$ML_{0.2}^{\mathrm{GLS}*}$	$ML_{0.4}^{\mathrm{GLS}*}$	$ML_{0.6}^{\mathrm{GLS}*}$
1	c=3	7.04	7.64	7.72	7.70	8.06	8.12
	c=6	13.88	14.80	14.88	14.38	15.06	14.96
	c=9	25.04	26.14	26.76	25.90	26.56	26.50
2	c=3	8.06	9.32	10.22	8.84	9.38	10.02
	c=6	14.80	16.54	17.72	15.96	16.90	17.70
	c=9	26.00	28.34	30.64	26.96	28.62	29.78
3	c=3	7.10	8.14	8.94	7.24	7.30	7.78
	c=6	13.32	15.32	16.68	14.04	14.50	15.18
	c=9	25.74	27.94	29.74	26.92	27.60	27.84
4	c=3	6.04	6.82	7.12	5.98	6.12	6.68
	c=6	11.52	12.90	13.72	11.86	12.16	12.82
	c=9	22.10	23.94	24.96	22.14	22.80	23.64
5	c=3	7.54	8.36	10.38	8.16	8.46	9.02
	c=6	14.52	16.02	18.62	15.46	15.56	16.62
	c=9	26.24	28.96	32.66	27.68	27.90	28.56
6	c=3	7.72	9.60	11.96	7.70	8.94	10.60
	c=6	13.96	16.90	19.34	14.44	15.98	18.04
	c=9	25.78	28.50	30.78	25.74	26.96	29.64
7	c=3	7.34	8.50	9.00	7.40	8.76	9.04
	c=6	13.92	15.76	16.56	14.24	15.82	16.94
	c=9	25.86	27.94	28.98	26.48	27.96	28.88
8	c=3	7.88	10.80	12.92	7.48	10.46	12.54
	c=6	14.28	18.70	21.12	14.66	18.32	20.92
	c=9	24.84	30.08	32.74	25.90	29.80	33.24
9	c=3	7.92	9.44	11.24	7.48	8.72	10.04
	c=6	14.22	16.86	19.06	14.30	15.86	17.50
	c=9	24.78	27.98	30.66	25.78	27.90	29.62
10	c=3	6.68	11.38	16.00	6.78	9.42	13.66
	c=6	11.04	17.52	23.78	11.66	15.08	20.96
	c=9	17.34	25.68	33.12	17.00	22.16	29.94

1.7 Conclusion

In this chapter we show that the heteroskadastic innovations may affect the asymptotic distribution of unit root tests. We consider three recent adjustments to correct this problem. We suggest a detrended version of the Phillips-Perron test by Beare (2006) and propose an extension of the ML test by Boswijk (2005) to allow for deterministic components. To investigate the finite sample properties of the unit root tests we perform a simulation study with two different specifications of the deterministic components of the data generating process. The key results of the simulation study can be summarized as follows. First, the proposed detrended variant of the Phillips-Perron by Beare (2006) clearly outperforms the conventional test by Beare (2006). Second, the choice of the bandwidth parameter h has a significant effect on the empirical size and the power of the tests. Third, the ML test is superior to the remaining tests in terms of power. However, one has to bear in mind that the ML test requires nonparametric estimation of the volatility function and simulation methods to obtain critical values. Therefore, the computational effort is much more time-consuming than for both the bootstrap M tests and, in particular, the detrended version of the test by Beare (2006).

Chapter 2

Estimation and Testing in a Three-Regime Vector Error Correction Model

2.1 Introduction

Since its methodical foundation by Engle and Granger (1987) cointegration is one of the most important research areas in applied as well as theoretical time series analysis. An appropriate framework for the analysis of cointegration has been proved to be the linear vector error correction model (VECM). Its mathematical and statistical theory has been developed amongst others by Engle and Granger (1987) and Johansen (1988, 1991). However, an essential shortcoming of the linear VECM is that deviations from the long-run equilibrium are corrected by the same strength being independent of the magnitude of the equilibrium error. As this is often a questionable assumption (e.g. the validity of the purchasing power parity is questionable due to transaction costs), vector error correction models with nonlinear adjustment have been suggested recently.

van Dijk and Franses (2000) suggest a smooth transition vector error correction model in which the strength of adjustment increases gradually as the equilibrium error gets larger. They consider interest rate series for the Netherlands consisting of one- and twelve month interbank rates. They find that the smooth transition vector error correction model captures the dynamics clearly better than the linear vector error correction model.

An alternative model, the threshold vector error correction model introduced by Balke and Fomby (1997) and Lo and Zivot (2001), has attracted much more attention than the smooth transition vector error correction model. In this model, the adjustment speed changes depending on the value of the error correction term or another exogenous variable. In particular, it is possible that the variables evolve independently if the deviation from the long-run equilibrium is small, but the variables become cointegrated if the deviation exceeds a threshold. Most of the theoretical work focuses on the two-regime model. Bec and Rahbek (2004) examine the stability of the two-regime model and show that Δy_t and $\beta^\top y_t$ are geometrically ergodic processes under suitable assumptions. Hansen and Seo (2002) provide a full statistical treatment in a two-regime vector error correction model. In particular, they construct a LM test for threshold effects and show that the asymptotic distribution depends on moments of the data set. Hence, bootstrap methods are necessary to obtain critical values. However, Hansen and Seo (2002) assume that it is known a priori that the system is cointegrated with a single cointegrating vector. Furthermore, Seo (2004) develops a test for cointegration in the presence of possible threshold effects. Recently, Gonzalo and Pitarakis (2006) consider a two-regime threshold vector error correction model without imposing rank restrictions on the long-run impact matrices. However, they assume that lagged dependent variables and constants do not enter the model. In this setting, they construct a Wald test against threshold effects and show that the asymptotic distribution is nuisance-free and independent of the absence or presence of unit roots and cointegration.

The present chapter contributes to this line of research. We consider a general three-regime threshold vector error correction model that does not impose any rank restrictions on the long-run impact matrices. We prove consistency of the threshold estimators obtained by minimizing the determinant of the estimated covariance matrix. An information-based selection procedure is introduced to estimate the cointegration ranks. A supLM test for linearity is suggested that takes advantage of the statistical properties of the process. It is shown that the asymptotic distribution depends on moments of the data set. Hence, we propose a parametric bootstrap method to obtain critical values. Finally, we apply the proposed econometric methodology to the term structure of interest rates. We find strong evidence for threshold effects. The results confirm the intuition that the series are not cointegrated in the case of small deviations from the long-run equilibrium, but that they become cointegrated for large deviations. Furthermore, the model clearly outperforms the random walk model and the linear error correction model in terms of forecast ability.

The chapter is organized as follows. In section 2.2 we introduce the three-regime threshold VECM and give different examples illustrating the flexibility of the model. Section 2.3 is devoted to the estimation of the threshold parameters. The problem of selecting an appropriate rank configuration is discussed in section 2.4. A supLM test against threshold effects is suggested in section 2.5. The results of the empirical application are presented in section 2.6. Finally, we conclude. All mathematical proofs are postponed to the appendix.

2.2 The Three-Regime Threshold VECM

In the following we consider the model

$$\begin{aligned}\Delta y_t &= \left(\mu_1 + \Pi_1 y_{t-1} + \sum_{i=1}^{k} \Gamma_{1,i} \Delta y_{t-i}\right) I(q_{t-1} \leq \gamma_1) \\ &+ \left(\mu_2 + \Pi_2 y_{t-1} + \sum_{i=1}^{k} \Gamma_{2,i} \Delta y_{t-i}\right) I(\gamma_1 < q_{t-1} \leq \gamma_2) \\ &+ \left(\mu_3 + \Pi_3 y_{t-1} + \sum_{i=1}^{k} \Gamma_{3,i} \Delta y_{t-i}\right) I(q_{t-1} > \gamma_2) + u_t, \qquad (2.1)\end{aligned}$$

where Δ is the difference operator, $(y_t)_{t=1,\dots,T} = \left((y_{1,t},\dots,y_{p,t})^\top\right)_{t=1,\dots,T}$ is a p-dimensional process and $I(\cdot)$ is the indicator function being 1 if the statement in brackets is true and being 0 else. The parameters μ_1,μ_2,μ_3 and $\Gamma_{l,1},\dots,\Gamma_{l,k}$, $l = 1,2,3$, are unknown $p\times 1$ and $p\times p$-matrices, respectively. The long-run impact matrices Π_1,Π_2 and Π_3 are possibly of reduced rank denoted by $0 \le \mathrm{Rank}(\Pi_l) = r_l \le p$. It is important to emphasize that the cointegrating vectors, if they exist, are assumed to be constant throughout all regimes. Therefore, Π_l is decomposed into $\Pi_l = \alpha_l\beta^\top$. The thresholds γ_1 and γ_2 are unknown real numbers with $\gamma_1 < \gamma_2$. Finally, $(u_t)_{t=1,\dots,T}$ denotes a p-dimensional error process. In order to illustrate the rich dynamics of model (2.1), we discuss two basic models. We exclude lagged dependent variables and constants to keep the examples simple.

Example 1:

We consider a bivariate system of cointegrated $I(1)$ variables. We set

$$\Pi_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \Pi_1 = \Pi_3 = \begin{pmatrix} -0.6 & 0.6 \\ 0.7 & -0.7 \end{pmatrix}$$

and $q_t = \beta^\top y_t = (1,-1)^\top y_t$. Since the rank of both Π_1 and Π_3 is one, we decompose them into $\Pi_l = \alpha\beta^\top$ $(l = 1,3)$ with $\alpha = (-0.6, 0.7)^\top$ and $\beta = (1,-1)^\top$. If $\beta^\top y_t \in (\gamma_1,\gamma_2]$, y_t and $\beta^\top y_t$ follow a random walk and, hence, the components of y_t are not cointegrated. Furthermore, if $\beta^\top y_t$ is smaller than γ_1 or larger than γ_2, $\beta^\top y_t$ follows the stationary $AR(1)$-process

$$\beta^\top y_t = -0.3\beta^\top y_{t-1} + \eta_t,$$

with $\eta_t = \beta^\top \epsilon_t$. Hence, the components of y_t are cointegrated.

Example 2:

Now we consider a bivariate system of stationary variables. We set

$$\Pi_i = \begin{pmatrix} \rho_{11}^{(l)} & 0 \\ 0 & \rho_{22}^{(i)} \end{pmatrix}, \qquad l = 1, 2, 3.$$

Hence, we have

$$\begin{aligned} y_{1,t} &= \delta_{11}^{(1)} y_{1,t-1} I(q_{t-1} <= \gamma_1) + \delta_{11}^{(2)} y_{1,t-1} I(\gamma_1 < q_{t-1} <= \gamma_2) + \delta_{11}^{(3)} y_{1,t-1} I(q_{t-1} > \gamma_2) + \epsilon_{1,t} \\ y_{2,t} &= \delta_{22}^{(1)} y_{2,t-1} I(q_{t-1} <= \gamma_1) + \delta_{22}^{(2)} y_{2,t-1} I(\gamma_1 < q_{t-1} <= \gamma_2) + \delta_{22}^{(3)} y_{2,t-1} I(q_{t-1} > \gamma_2) + \epsilon_{2,t}, \end{aligned}$$

with $\delta_{jj}^{(l)} = \rho_{jj}^{(l)} + 1$, $l = 1, 2, 3$ and $j = 1, 2$. According to Theorem 4.2 in Fan and Yao (2003), $(y_{j,t})_{t=1,\ldots,T}$ $(j = 1, 2)$ is stationary if $\max_{1 \le i \le 3} |\delta_{jj}^{(i)}| < 1$.

For convenience, we re-write (2.1) by

$$\Delta Y = A_1 Z_1 + A_2 Z_2 + A_3 Z_3 + U, \tag{2.2}$$

where A_l, $l = 1, 2, 3$, denotes a matrix of the form

$$A_l = (\mu_l, \Pi_l, \Gamma_{l,1}, \ldots, \Gamma_{l,k})$$

and the matrices ΔY, Z_l, $l = 1, 2, 3$, and U are defined as

$$\begin{aligned} \Delta Y &= (\Delta y_1, \ldots, \Delta y_T), \\ Z_l(\gamma_{l-1}, \gamma_l) &= (\xi_0 I(\gamma_{l-1} < q_0 \le \gamma_l), \ldots, \xi_{T-1} I(\gamma_{l-1} < q_{T-1} \le \gamma_l)), \\ U &= (u_1, \ldots, u_T), \end{aligned}$$

where

$$\xi_{t-1} = \left(1, y_{t-1}^\top, \Delta y_{t-1}^\top, \ldots, \Delta y_{t-k}^\top\right)^\top$$

is a $p \cdot m$-dimensional vector with $m = (k+1)p+1$, and we set $\gamma_0 = -\infty$ and $\gamma_3 = \infty$. Note that Z_i and Z_j $(i, j = 1, 2, 3; i \neq j)$ are orthogonal by definition, i.e. $Z_i Z_j^\top = 0$ $(i \neq j)$, and $Z = (\xi_0, \ldots, \xi_{T-1}) = Z_1(\gamma_1) + Z_2(\gamma_1, \gamma_2) + Z_3(\gamma_3)$. In the following we suppress their dependence on γ_1 and γ_2, i.e. $Z_1 = Z_1(\gamma_1)$, $Z_2 = Z_2(\gamma_1, \gamma_2)$ and $Z_3 = Z_3(\gamma_2)$.

Under the assumption that the errors u_t are i.i.d. Gaussian $(0, \Omega)$, we suggest estimation of model (2.2) by maximum likelihood. The log-likelihood function is given by

$$\begin{aligned} \mathcal{L}(A_1, A_2, A_3, \Omega, \gamma_1, \gamma_2) &= -\frac{Tp}{2} \ln(2\pi) - \frac{T}{2} \ln |\Omega| - \frac{1}{2} \sum_{t=1}^{T} u_t^\top \Omega^{-1} u_t \\ &= -\frac{Tp}{2} \ln(2\pi) - \frac{T}{2} \ln |\Omega| \\ &- \frac{1}{2} tr \left\{ (\Delta Y - A_1 Z_1 - A_2 Z_2 - A_3 Z_3)^\top \Omega^{-1} \right. \\ &\cdot \left. (\Delta Y - A_1 Z_1 - A_2 Z_2 - A_3 Z_3) \right\}. \end{aligned} \tag{2.3}$$

For given γ_1 and γ_2, (2.3) is maximized by the estimators

$$\hat{A}_1(\gamma_1) = \Delta Y Z_1^\top (Z_1 Z_1^\top)^{-1}, \tag{2.4}$$

$$\hat{A}_2(\gamma_1, \gamma_2) = \Delta Y Z_2^\top (Z_2 Z_2^\top)^{-1}, \tag{2.5}$$

$$\hat{A}_3(\gamma_2) = \Delta Y Z_3^\top (Z_3 Z_3^\top)^{-1}, \tag{2.6}$$

$$\hat{\Omega}(\gamma_1, \gamma_2) = \frac{1}{T} \hat{U}(\gamma_1, \gamma_2) \hat{U}(\gamma_1, \gamma_2) \top, \tag{2.7}$$

with $\hat{U}(\gamma_1, \gamma_2) = \Delta Y - \hat{A}_1(\gamma_1) Z_1 - \hat{A}_2(\gamma_1, \gamma_2) Z_2 - \hat{A}_3(\gamma_2) Z_3$. Note that the estimators (2.4)-(2.7) coincide with the estimators obtained by OLS. For further use, we define the vectorised versions of (2.4), (2.5) and (2.6), i.e.

$$\text{vec } \hat{A}_1 = [(Z_1 Z_1^\top)^{-1} Z_1 \otimes I_p] \text{vec } \Delta Y, \tag{2.8}$$

$$\text{vec } \hat{A}_2 = [(Z_2 Z_2^\top)^{-1} Z_2 \otimes I_p] \text{vec } \Delta Y, \tag{2.9}$$

$$\text{vec } \hat{A}_3 = [(Z_3 Z_3^\top)^{-1} Z_3 \otimes I_p] \text{vec } \Delta Y, \tag{2.10}$$

where $\otimes$ is the Kronecker product operator.

2.3 Estimation of the threshold parameters

We start by suggesting an estimation procedure of the thresholds γ_1 and γ_2. According to the least square approach, the estimators of the threshold parameters are obtained by minimizing the determinant of the covariance matrix $\hat{\Omega}(\gamma_1, \gamma_2)$ being defined by (2.7), i.e.

$$(\hat{\gamma}_1, \hat{\gamma}_2) = \operatorname{argmin}_{\gamma_1, \gamma_2 \in \Gamma} \left| \hat{U}(\gamma_1, \gamma_2) \hat{U}(\gamma_1, \gamma_2)^\top \right|. \tag{2.11}$$

For the subsequent analysis we make the following assumptions

(A1) The process $w_t := (\Delta y_t, q_t)$ is L^{4r}-bounded, strictly stationary and β-mixing with mixing coefficient $\beta_m = O(m^{-A})$ where $A > \nu/(\nu - 1)$ and $r > \nu > 1$.

(A2) The process $(u_t)_{t=1,\dots,T}$ is a p-dimensional vector white noise sequence satisfying $E(u_t) = 0$ and $E(u_t u_t^\top) = \Omega$, where Ω is a symmetric and positive definite matrix.

(A3) The process $(q_t)_{t=1,\dots,T}$ is strictly stationary, ergodic and independent of the process $(u_t)_{t=1,\dots,T}$. We assume that the process $(q_t)_{t=1,\dots,T}$ has a distribution F being continuous everywhere. Furthermore, we set $r = F(\gamma)$ such that $I(q_{t-1} \leq \gamma) = I(\phi_{t-1} \leq r)$ where $\phi_{t-1} = F(q_{t-1})$.

(A4) The thresholds γ_1 and γ_2 are contained in $\Gamma = [\gamma_L, \gamma_U]$, a closed and bounded subset of $\mathbb{R}$.

Recall that the β-mixing coefficient β_m is defined as

$$\beta_m = \sup_t E \left[\sup_{A \in \mathcal{F}_{t+m}^{\infty}} |P(A) - P(A|\mathcal{F}_{\infty}^{t})| \right],$$

where $\mathcal{F}_s^t$ denotes the σ-algebra generated by $(w_s, \ldots, w_t)$ for $s \leq t$.

It is well-known that the β-mixing condition in Condition (A1) holds for many processes such as stationary ARMA processes and ARCH models, see Mokkadem (1988) and Masry and Tjostheim (1995), respectively. Note that a strong (weak) moment restriction on w_t leads to weak (strong) condition on the convergence speed of β_m.

Condition (A2) ensures that appropriate central limit theorems are applicable. However, it may be relaxed to the assumption of a martingale difference sequence.

Condition (A3) ensures that the process $(q_t)_{t=1,\ldots,T}$ is well-behaved and to exclude that it is $I(1)$ itself.

Finally, condition (A4) is a standard assumption. Following Andrews (1993), suitable choices of γ_L and γ_U satisfy $P(q_{t-1} \leq \gamma_L) > \theta$, $P(\gamma_L < q_{t-1} \leq \gamma_U) > \theta$ and $P(q_{t-1} > \gamma_U) > \theta$ with θ usually chosen to be 0.1 or 0.15.

Under these assumptions, the following proposition establishes the consistency of $\hat{\gamma}_1$ and $\hat{\gamma}_2$. In particular, we have

Theorem 2.1:
Under Assumptions (A1)-(A4), $\hat{\gamma}_1$ and $\hat{\gamma}_2$ satisfying (2.11) are consistent estimators for γ_1 and γ_2, i.e. $\hat{\gamma}_1 \rightarrow_p \gamma_1$ and $\hat{\gamma}_2 \rightarrow_p \gamma_2$ as $T \rightarrow \infty$.

In order to examine the finite-sample behaviour of the estimators, we perform a simulation study. In the following we consider a stationary process and a cointegrated process. In both configurations we exclude the constant and lags of Δy_t and set the threshold parameters to $\gamma_1 = -0.5$ and $\gamma_2 = 0.5$. The threshold variable q_t is taken to be a standard normal iid random variable. The simulation experiments are conducted for both $T = 200$ and $T = 300$. The number of replications is chosen to be $N = 1000$. Furthermore, we set $\theta = 0.15$. The empirical means and the standard deviations of both estimators are presented in Table 2.1. We note that the estimation procedure

Table 2.1: Empirical means and standard deviations of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ (joint estimation)

	$\gamma_1 = -0.5$		$\gamma_2 = 0.5$	
	Mean($\hat{\gamma}_1$)	Std($\hat{\gamma}_1$)	Mean($\hat{\gamma}_2$)	Std($\hat{\gamma}_2$)
Cointegrated DGP				
T=200	-0.5104	0.0524	0.4902	0.0462
T=300	-0.5084	0.0330	0.4948	0.0327
Stationary DGP				
T=200	-0.5047	0.1093	0.4881	0.1073
T=300	-0.5050	0.0626	0.4969	0.0626

works very well under both configurations. The bias of the estimators as well as the standard deviation decline with the sample size. Since the computation time [1] increases with the sample size at high rate, Bai (1997) suggests a sequential estimation procedure. Following Bai (1997), the estimation procedure involves two steps. On the first stage a misspecified two-regime threshold model is estimated. Bai (1997) shows that the obtained estimator $\hat{\gamma}$ is consistent for one of the two thresholds (γ_1, γ_2). At a second stage, a three-regime threshold model is estimated under the constraint that one element of (γ_1, γ_2) equals $\hat{\gamma}$. In order to see the enormous computational savings of this procedure, let us denote the number of evaluations by N. Then it is obvious that instead of N^2 evaluations only $N + 2N$ evaluations are involved. The results using the sequential estimation procedure are presented in Table 2.2. Comparing the empirical means of both estimation procedures we note that the latter suffer from a higher, but still moderate bias. Furthermore, the estimates obtained by using the sequential procedure are more volatile demonstrated by higher standard deviations. However, the sequential procedure is predominantly used in practice due to the enormous computational savings [2].

[1] The computation time is about 1.5 hours when $T = 200$ and about eight hours when $T = 300$.

[2] The computation time is about 40 seconds when $T = 200$ and about two minutes when $T = 300$.

Table 2.2: Empirical means and standard deviations of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ (seq. estimation)

	$\gamma_1 = -0.5$		$\gamma_2 = 0.5$	
	Mean($\hat{\gamma}_1$)	Std($\hat{\gamma}_1$)	Mean($\hat{\gamma}_2$)	Std($\hat{\gamma}_2$)
Cointegrated DGP				
T=200	-0.5252	0.0645	0.5045	0.0684
T=300	-0.5204	0.0490	0.5061	0.0470
Stationary DGP				
T=200	-0.5411	0.1164	0.5158	0.1259
T=300	-0.5249	0.0769	0.5179	0.0839

2.4 Estimation of the cointegration rank

Having estimated the threshold parameters the next step involves the determination of the cointegration ranks of the long-run impact matrices Π_1, Π_2 and Π_3. Following Phillips and Chao (1999) and Gonzalo and Pitarakis (2006) we use an information-based selection approach. In the spirit of the usual information criterions (see e.g. Lütkepohl, 2006) we optimize an objective function balancing the goodness-of-fit and the number of estimated parameters. We define the matrices

$$\begin{aligned}
\widehat{\Delta Y_l} &= (\Delta y_1 I(\hat{\gamma}_{l-1} < q_0 \leq \hat{\gamma}_l), \ldots, \Delta y_T I(\hat{\gamma}_{l-1} < q_{T-1} \leq \hat{\gamma}_l)), \\
\widehat{Y_{l,-1}} &= (y_0 I(\hat{\gamma}_{l-1} < q_0 \leq \hat{\gamma}_l), \ldots, y_{T-1} I(\hat{\gamma}_{l-1} < q_{T-1} \leq \hat{\gamma}_l)), \\
\hat{Z}_l &= (\xi_0 I(\hat{\gamma}_{l-1} < q_0 \leq \hat{\gamma}_l), \ldots, \xi_{T-1} I(\hat{\gamma}_{l-1} < q_{T-1} \leq \hat{\gamma}_l)), \\
\widehat{Z_{l,-1}} &= (\zeta_0 I(\hat{\gamma}_{l-1} < q_0 \leq \hat{\gamma}_l), \ldots, \zeta_{T-1} I(\hat{\gamma}_{l-1} < q_{T-1} \leq \hat{\gamma}_l)), \\
\zeta_{t-1} &= (1, \Delta y_{t-1}^\top, \ldots, \Delta y_{t-k-1}^\top)^\top, \\
\hat{U}_l &= \widehat{\Delta Y_l} - \hat{A}_l \hat{Z}_l, \\
\hat{\Omega}_l &= T^{-1} \hat{U}_l \hat{U}_l^\top
\end{aligned}$$

with $l = 1, 2, 3$ and $\hat{\gamma}_0 = -\infty$ and $\hat{\gamma}_3 = \infty$. Regressing $\widehat{\Delta Y_l}$ and $\widehat{Y_{l,-1}}$ on $\widehat{Z_{l,-1}}$, we obtain the residuals $R_{l,0}$ and $R_{l,1}$, respectively. Next, we introduce the product moment matrices

$$\begin{aligned} S_{l,00} &= T^{-1} R_{l,0} R_{l,0}^{\top}, \\ S_{l,11} &= T^{-1} R_{l,1} R_{l,1}^{\top}, \\ S_{l,01} &= T^{-1} R_{l,0} R_{l,1}^{\top}, \\ S_{l,10} &= S_{l,01}^{\top}, \end{aligned}$$

with $l = 1, 2, 3$. Note that $\hat{\Omega}_l = S_{l,00} - S_{l,01} \left(S_{l,11}\right)^{-1} S_{l,10}$ by direct calculation. In the following $\hat{\Omega}_l(r_l)$ denotes the sample covariance matrix of regime l assuming rank $(\Pi_l) = r_l$. Our estimator is defined as

$$\hat{r}_l = \text{argmin}_{r_l \in \{0,\ldots,p\}} IC_l^*(r_l),$$

where

$$IC_l^*(r_l) = \ln \left| \hat{\Omega}_l(r_l) \right| + \frac{c_T}{T} m(r_l),$$

with $m(r_l)$ denotes the number of estimated parameters and c_T is a deterministic penalty term depending on T. Using standard arguments we get

$$\ln \left| \hat{\Omega}_l(r_l) \right| = \ln \left(S_{l,00}\right) + \sum_{i=1}^{r_l} \ln \left(1 - \lambda_{l,i}\right), \tag{2.12}$$

where $\lambda_{l,i}$ denotes the i-th eigenvalue of $\left(S_{l,00}\right)^{-1} S_{l,01} \left(S_{l,11}\right)^{-1} S_{l,10}$. Since $S_{l,00}$ is independent of the value of r_l, it is sufficient to optimize

$$IC_l(r_l) = \sum_{i=1}^{r_l} \ln \left(1 - \lambda_{l,i}\right) + \frac{c_T}{T}(2pr_l - r_l^2),$$

with $1 > \lambda_{l,1} > \cdots > \lambda_{l,p}$. The following proposition states the asymptotic properties of the described model selection approach. We omit its proof since it follows exactly the same steps as the proof of Proposition 6 in Gonzalo and Pitarakis (2006).

Theorem 2.2:
Under the assumptions (i) $c_T \to \infty$ and (ii) $\frac{c_T}{T} \to 0$, the estimator $\hat{r}_l$ defined as in (2.12) converges in probability to its true value as $T \to \infty$, i.e. $\hat{r}_l \to_p r_l$, where r_l denotes the true rank of Π_l, $l = 1, 2, 3$.

In the following simulation study and the empirical application we use $c_T = \ln T$ satisfying both (i) and (ii). This corresponds to the well-known Bayesian Information Criterion (BIC).

In order to evaluate the finite sample properties of the model selection approach, we perform a simulation study. The data are generated according to three different processes; the rank configuration of DGP I is $r_1 = 1$, $r_2 = 0$ and $r_3 = 1$, that of DGP II is $r_1 = 1$, $r_2 = 1$ and $r_3 = 1$, and that of DGP III is $r_1 = 2$, $r_2 = 2$ and $r_3 = 2$. Throughout all DGPs the threshold variable is a standard normally distributed random variable, the threshold parameters are set to $\gamma_1 = -0.5$ and $\gamma_2 = 0.5$. These choices ensure that the number of observations is approximately equal in each regime. The results based on $N = 3000$ replications are shown in Table 2.3.

The data indicate good finite sample properties of the model selection approach. The decision frequency reaches a low of 46.20 % for DGP III when $T = 100$, of 59.37 % for DGP III when $T = 300$ and of 75.10 % when for DGP II when $T = 500$.

2.5 A test against threshold effects

It is obvious that regimes featuring *different* ranks are subject to threshold effects. The aim of this section is to construct a test that enables us to detect threshold effects between regimes which are characterized by the *same* rank $0 \leq r \leq p$. Since the need for estimating the cointegrating vectors only occurs when $0 < r < p - 1$, we discuss this situation in detail. Extensions covering the situations $r = 0$ and $r = p$ are provided at the end of the section.

Table 2.3: Decision frequencies of the model selection approach, $c_T = \ln T$

	$\hat{r}_1 = 0$	$\hat{r}_1 = 1$	$\hat{r}_1 = 2$	$\hat{r}_2 = 0$	$\hat{r}_2 = 1$	$\hat{r}_2 = 2$	$\hat{r}_3 = 0$	$\hat{r}_3 = 1$	$\hat{r}_3 = 2$
				DGP I: $r_1 = 1, r_2 = 0, r_3 = 1$					
T=100	1.13	63.74	35.13	53.07	33.27	13.66	0.93	63.24	35.83
T=300	0.00	75.00	25.00	80.10	17.23	2.67	0.00	78.03	21.97
T=500	0.00	79.43	20.57	86.20	12.30	1.50	0.00	81.77	18.23
				DGP II: $r_1 = 1, r_2 = 1, r_3 = 1$					
T=100	1.10	61.27	37.63	15.33	53.77	30.90	1.40	60.80	37.80
T=300	0.00	69.70	30.30	8.07	73.40	18.53	0.00	72.27	27.73
T=500	0.00	75.10	24.90	1.27	83.27	15.46	0.00	77.63	22.37
				DGP III: $r_1 = 2, r_2 = 2, r_3 = 2$					
T=100	1.93	20.07	78.00	15.13	38.67	46.20	2.13	19.70	78.17
T=300	0.00	1.47	98.53	12.27	28.36	59.37	0.03	1.87	98.10
T=500	0.00	0.10	99.90	5.00	18.60	76.40	0.00	0.33	99.67

2.5.1 Test statistic

Using the rank restriction $0 < r < p$, model (2.2) can be re-written as

$$\Delta Y = A_\alpha Z_{\hat\beta} + \delta_{\alpha_1} Z_{1,\hat\beta} + \delta_{\alpha_3} Z_{3,\hat\beta} + U,$$

with

$$\begin{aligned}
A_{l,\alpha} &= (\mu, \alpha_l, \Gamma_{l,1}, \ldots, \Gamma_{l,k}) \\
Z_{\hat\beta} &= \left(\xi_{0,\hat\beta}, \ldots, \xi_{T-1,\hat\beta}\right), \\
Z_{l,\hat\beta} &= \left(\xi_{0,\hat\beta} I(\gamma_{l-1} < q_0 \le \gamma_l), \ldots, \xi_{T-1,\hat\beta} I(\gamma_{l-1} < q_{T-1} \le \gamma_l)\right), \\
\xi_{t-1,\hat\beta} &= \left(1, y_{t-1}^\top \hat\beta, \Delta y_{t-1}^\top, \ldots, \Delta y_{t-k}^\top\right)^\top
\end{aligned}$$

and $A_\alpha = A_{2,\alpha}$, $\delta_{1,\alpha} = A_{1,\alpha} - A_\alpha$ and $\delta_{3,\alpha} = A_{3,\alpha} - A_\alpha$. Hence, the hypotheses we are interested in can be expressed as

$$\begin{aligned}
H_0 &: \text{ vec } \delta_{1,\alpha} = \text{vec } \delta_{3,\alpha} = 0 \text{ against} \\
H_1 &: \neg H_0.
\end{aligned}$$

The standard expression for the LM-statistic is given by

$$\mathrm{LM}(\gamma_1, \gamma_2) = \text{vec}\begin{pmatrix} \hat\delta_{1,\alpha} \\ \hat\delta_{3,\alpha} \end{pmatrix}^\top \left[M_T(\gamma_1, \gamma_2)^{-1} \otimes \tilde\Omega^{-1}\right] \text{ vec}\begin{pmatrix} \hat\delta_{1,\alpha} \\ \hat\delta_{3,\alpha} \end{pmatrix}, \tag{2.13}$$

where $M_T(\gamma_1, \gamma_2)$ denotes a covariance estimator of $\text{vec}(\hat\delta_{1,\alpha}^\top, \hat\delta_{3,\alpha}^\top)^\top$. In particular, the estimator is of the form

$$M_T(\gamma_1, \gamma_2) = \begin{pmatrix} (Z_{1,\hat\beta} Z_{1,\hat\beta}^\top)^{-1} + (Z_{2,\hat\beta} Z_{2,\hat\beta}^\top)^{-1} & (Z_{2,\hat\beta} Z_{2,\hat\beta}^\top)^{-1} \\ (Z_{2,\hat\beta} Z_{2,\hat\beta}^\top)^{-1} & (Z_{2,\hat\beta} Z_{2,\hat\beta}^\top)^{-1} + (Z_{3,\hat\beta} Z_{3,\hat\beta}^\top)^{-1} \end{pmatrix}$$

and $\tilde\Omega = \frac{\tilde U \tilde U^\top}{T}$ is an estimator for Ω using the residuals under the null hypothesis. If the threshold parameters were known, then (2.13) would be asymptotically

$\chi^2(2p + 2pr + 2p^2k)$ distributed. Since we will not assume that the thresholds are known, standard distribution theory is not applicable. In particular, the so-called "Davies problem" occurs namely when the thresholds γ_1 and γ_2 are not identified under the null hypothesis. Here, we follow Davies (1987) by taking the supremum of (2.13) with respect to the thresholds. Although Andrews and Ploberger (1994) argue that the power of the test can be improved by using exponentially weighted averages of (2.13), we restrict our analysis to taking the supremum. It is important to note that the values of γ_1 and γ_2 which maximize the test statistic (2.13) will be in general different from $\hat{\gamma}_1$ and $\hat{\gamma}_2$ presented in section 2.3. The reason for this is that (2.13) is an LM test, and is therefore based on parameter estimates obtained under the null hypothesis rather than the alternative. The asymptotic distribution is given in the following theorem.

Theorem 2.3:
Suppose Assumptions (A1)-(A4) hold. Then as $T \to \infty$

$$\text{supLM} \Rightarrow \sup_{\theta \le r_1, r_2 \le 1-\theta} J(r_1, r_2)^\top \left[M(r_1, r_2) \otimes \Omega^{-1}\right] J(r_1, r_2), \tag{2.14}$$

where $J(r_1, r_2)$ and $M(r_1, r_2)$ are defined in the appendix.

It is important to emphasize that both $J(r_1, r_2)$ and $M(r_1, r_2)$ depend on $E[1(r_{l-1} < \phi_{t-1} \le r_l)\xi_{t-1,\beta}\xi_{t-1,\beta}^\top]$, $l = 1, 2, 3$; this means that the asymptotic distribution depends on moments of the data set and hence critical values have to be simulated for each data set. Here, we use the parametric residual bootstrap method. The steps of computing the bootstrap distribution are given as follows.

Step 1: Estimate the linear model given the rank restriction $\text{rk}(\Pi) = r$, yielding estimates $\hat{A}_\alpha^*$ and empirical residuals $\tilde{u}_t^*$.

Step 2: Draw from the residual vectors $\tilde{u}_t^*$ and create the vector series Δy_t^* and $\xi_{t-1,\beta}^*$ by recursion given $\hat{A}_\alpha^*$.

Step 3: Compute the test statistic supLM^* using Δy_t^* and $\xi_{t-1,\beta}^*$.

Step 4: Repeat Steps 1-3 B times and obtain the empirical distribution of the B test statistics of supLM.

Step 5: Let $\text{supLM}^*_{1-\alpha}$ be the $1-\alpha$ percentile of the empirical distribution obtained in the previous step. Reject the null hypothesis if $\text{supLM} > \text{supLM}^*_{1-\alpha}$.

We now turn to the cases $r = 0$ and $r = p$. If $r = 0$, we have

$$\xi_{t-1} = \left(1, \Delta y_{t-1}^\top, \ldots, \Delta y_{t-k}^\top\right)^\top$$

and if $r = p$, we have

$$\xi_{t-1} = \left(1, y_{t-1}, \Delta y_{t-1}^\top, \ldots, \Delta y_{t-k}^\top\right)^\top$$

It is obvious that ξ_{t-1} contains only stationary variables in both cases. Therefore, the previous analysis can be used to cover both situations.

In order to evaluate the empirical size of the proposed test we perform a small simulation study. The data are generated according to three different bivariate processes, a stationary system, a non-stationary system without cointegration and a cointegrated system. The long-run matrix of the stationary system is of rank two, that of the non-stationary system of rank zero and that of the cointegrated system of rank one. In all three cases, we include one lag. To achieve an accurate approximation of the bootstrap distribution, we set $B = 500$. We perform $N = 1000$ replications with sample size $T = 50$. The results of the simulation study are shown in Table 2.4. The results show that the empirical size of the supLM test is very close to the nominal size, even for the small sample size $T = 50$.

2.5.2 The case without lags of Δy_t and intercept

In this subsection we consider the important case that lags of Δy_t and intercept are not included in model (2.1), that is $\Gamma_{l,i} = 0$ and $\mu_l = 0$, $l = 1, 2, 3$ and $i = 1, \ldots, k$. In

Table 2.4: Empirical size (in percent) of the supLM test, $\theta = 0.15$

	$T = 50$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
Stationary DGP	10.90	5.30	1.50
Non-stationary DGP	10.60	4.30	0.70
Cointegated DGP	10.10	5.70	1.40

this situation, we are able to simulate critical values being valid independent of the data set. The following simulation aims to provide critical values.

Throughout all our experiments the error process u_t is two-dimensional standard normally distributed with covariance matrix I_2. Furthermore, the threshold variable q_t follows a normally distributed white noise process. We perform 10000 replications with a sample size of $T = 400$. The results are presented for $\theta = 0.1$ and $\theta = 0.15$ in Table 2.5 and Table 2.6, respectively.

Table 2.5: Critical values of the supLM test (p=2, $\theta = 0.1$)

	90%	95%	97.5%	99%
Stationary DGP	27.715	30.513	32.912	36.572
Nonstationary DGP	27.697	30.472	32.724	36.424
Cointegrated DGP	27.685	30.598	32.906	36.611

Table 2.6: Critical values of the supLM test (p=2, $\theta = 0.15$)

	90%	95%	97.5%	99%
Stationary DGP	25.255	27.866	30.264	33.427
Nonstationary DGP	25.333	27.989	30.393	33.322
Cointegrated DGP	25.122	27.647	30.184	33.171

The results in Table 2.5 and 2.6 clearly show that the critical values are very similar independent of the stochastic properties of the VECM. This enables us to test for threshold effects without having knowledge about the absence or presence of unit roots and cointegration.

2.6 Application to the term structure of interest rates

In this section we apply the proposed econometric methodology to the term structure of interest rates. Due to the existence of transaction costs linear adjustment is not implied by the theory of the term structure.

In section 2.6.1 we briefly re-state the arguments by Campbell and Shiller (1987) that interest rates on bonds with different maturities are cointegrated. Section 2.6.2 is devoted to the description of the data. The empirical results are presented in section 2.6.3.

2.6.1 Expectations hypothesis

The expectations hypothesis asserts that the long-term interest rate R_t with a maturity of N is determined by the average of current and expected future returns on the bond with a maturity of one period, denoted by r_t. Hence, we obtain

$$R_t = \frac{1}{N}\sum_{i=1}^{N} E_t(r_{t+i-1}) = \frac{1}{N}\left[\sum_{i=1}^{N-1} E_t(r_{t+i}) + r_t\right]. \tag{2.15}$$

Obviously, it holds that

$$E_t(r_{t+i}) = \sum_{j=1}^{i} E_t(\Delta r_{t+j}) + r_t. \tag{2.16}$$

Combining (2.15) and (2.16) leads to

$$s_t = R_t - r_t = \frac{1}{N}\sum_{i=1}^{N-1}\sum_{j=1}^{i} E_t(\Delta r_{t+j}).$$

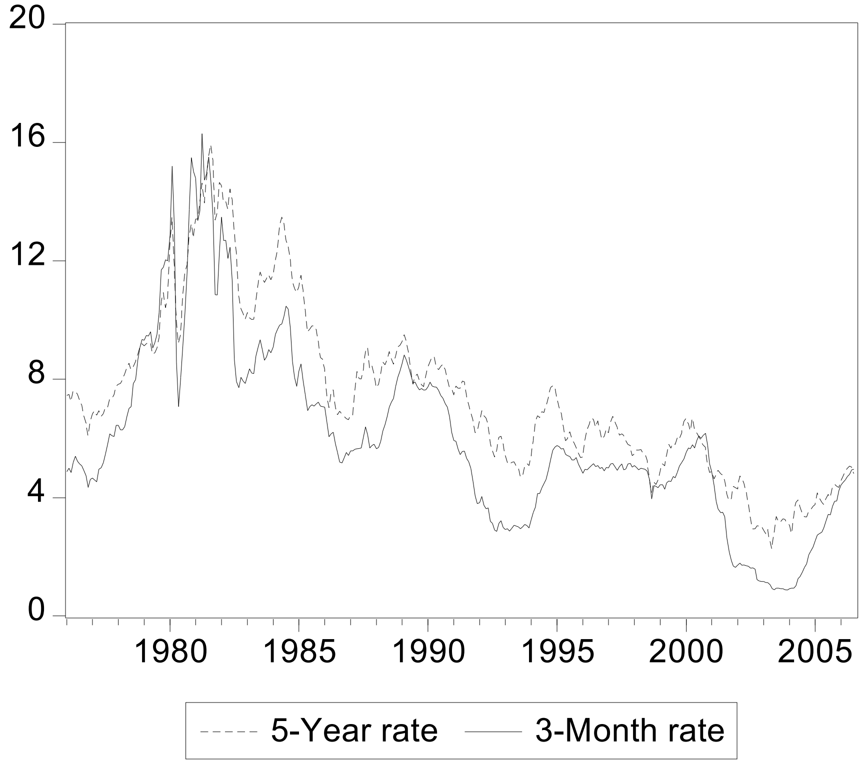

Figure 2.1: 5-Year interest rate and 3-Month interest rate (Jan. 1976 to Aug. 2006)

If we assume that the expected future change of the short-term interest rate is stationary, the term spread s_t is stationary. According to the definition of cointegration, the long-term interest rate R_t and the short-term interest rate r_t are cointegrated by the cointegrating vector $\beta^\top = (1, -1)^\top$.

2.6.2 Data

In our empirical application we use the monthly interest rate on the 3-month treasury bill as the short-term rate and the 5-year treasury constant maturity rate as the long-term rate. The sample period runs from January 1976 to August 2006. The data is extracted from the St. Louis federal reserve data base. Both series are depicted in Figure 2.1.

2.6.3 Empirical results

We start our analysis with the basic linear error correction model. The VAR lag-length selection is carried out by using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Table 2.7 shows the results for the range $k = 1$ to $k = 5$.

Table 2.7: Lag-length selection using AIC and BIC

	AIC	BIC
k=1	-744.31	-739.80
k=2	-761.27*	-754.52*
k=3	-759.32	-750.34
k=4	-756.13	-744.93
k=5	-749.91	-736.50

Table 2.7 shows that both Information Criteria suggest to choose $k = 2$. Next, we perform Johansen's (1988) cointegration test to determine the number of cointegrating relationships. Since we include an unrestricted constant in the model, the critical values of the test are chosen according to Table 15.3 in Johansen (1995). The results are given in the following table.

Table 2.8: Results of the trace test by Johansen (1995)

Null hypothesis	Test statistic	5% critical value
Rank=0	18.63	15.34
Rank=1	2.34	3.84

The results shown in Table 2.8 clearly indicate that a long-run relationship exists between the short-term interest rate and the long-term interest rate. The estimation results of the linear error correction model are presented in Table 2.9. Standard errors are based on the heteroskedasticity-robust covariance estimator. Due to space constraints we do not report the results of the lagged dependent variables. The

Table 2.9: Estimation results of the linear VECM (Jan. 1976 - Aug. 2006)

	Long-term interest rate	Short-term interest rate
β		-1.011
α	-0.032	0.035
s.e.(α)	0.024	0.040

cointegrating vector is estimated to be close to 1. The adjustment coefficient of the short-term rate is positive but not significant and hence, the long-run relationship does not provide significant information on the future change in the short rate. This finding is consistent with the results of Mankiw and Miron (1986). The adjustment coefficient of the long-term rate is not significant as well.

Due to the existence of transaction costs we conjecture that a threshold error correction model may provide a better empirical description. The estimates of the threshold parameters are $\hat{\gamma}_1 = 0.090$ and $\hat{\gamma}_2 = 2.430$. Next, we estimate the cointgration rank r_l of regime l $(l = 1, 2, 3)$. Using the information-based selection approach we obtain the results presented in Table 2.10.

Table 2.10: Choice of the cointegration rank

Rank	Regime 1	Regime 2	Regime 3
λ_1	0.452	0.030	0.345
λ_2	0.016	0.002	0.011
IC(r=0)	0	0*	0
IC(r=1)	-0.553 *	0.017	-0.376 *
IC(r=2)	-0.552	0.031	-0.371

The eigenvalue λ_1 is substantially different for the outer regimes, whereas λ_1 is close to zero for the middle regime. The eigenvalue λ_2 is close to zero for all regimes. Hence, the information-based approach suggests a cointegration rank of one for regime 1 and regime 3, and a cointegration rank of zero for regime 2. Given these rank restrictions we estimate a threshold error correction model whose estimation results are displayed in Table 2.11. Standard errors are calculated from the heteroskedasticity-robust covariance estimator. The trimming parameter θ is set at 0.1.

Table 2.11: Estimation results of the threshold VECM (Jan. 1976 - Aug. 2006)

	Long-term interest rate	Short-term interest rate
β		-1.025
α_1	-0.074	0.578
s.e.(α_1)	0.216	0.248
α_3	-0.650	-0.193
s.e.(α_3)	0.175	0.228
$\gamma = (\gamma_1, \gamma_2)$	0.090, 2.430	

Regime 1 contains 10.1% of the observations and covers the case that the difference between the short-term rate and the long-term rate is low. The short-term rate responds by a strong positive adjustment whereas the long-term rate does not respond reflected by a negative but insignificant adjustment coefficient. Regime 3 contains 17.8% of the observations. The adjustment coefficient of the short-term rate is negative, but insignificant. The long-term rate responds by a strong negative and significant adjustment coefficient.

In order to compare the forecasting ability of the threshold model with that of the random walk model and the linear model, we calculate the root mean squared error (RMSE) and mean absolute error (MAE) for all three models. The RMSE and the MAE are defined as

$$
\begin{aligned}
RMSE &= \sqrt{1/(T-k)\sum_{t=k}^{T}(\hat{E}_{t-1}r_t - r_t)^2}, \\
MAE &= 1/(T-k)\sum_{t=k}^{T}|\hat{E}_{t-1}r_t - r_t|.
\end{aligned}
$$

We set $k = 50$ to ensure that the parameter estimates are based on a sufficiently

Table 2.12: Comparison of the prediction ability of the random walk model, the linear VECM and the threshold VECM

	Random walk (A)	Linear VECM (B, B/A)	Threshold VECM (C, C/A)
RMSE	0.522	0.424 (0.812)	0.330 (0.632)
MAE	0.264	0.256 (0.970)	0.208 (0.788)

large number of observations. The results are shown in Table 2.12. The root mean squared error (RMSE) of the threshold VECM is about 35% lower than the random walk model while the linear VECM reduces the RMSE by about 20%. The mean absolute error (MAE) of the linear model is similar to the random walk model while the threshold VECM reduces the MAE by about 20%. Thus, the threshold VECM clearly improves the prediction ability.

2.7 Conclusion

In this chapter we examine a three-regime threshold error correction model being widely used in applications. In contradiction to previous contributions we do not make assumptions involving the stochastic properties of the process. In this general framework we propose to estimate the threshold parameters by minimizing the determinant of the residual covariance matrix and establish the consistency of the estimators. Since a joint optimization is computationally burdensome we introduce a sequential estimation procedure similar to Bai (1997). We propose an information-based approach balancing the goodness of fit and the number of estimated parameters to determine the rank configuration of the regimes. For the situation that the regimes are characterized by the *same* rank, we propose a supLM test to detect threshold effects. It turns out that the asymptotic distribution depends on moments of the data set and, hence, bootstrap methods are necessary to obtain critical values. In a simulation study it appears that the test shows low size distortion, even for small sample sizes. Finally, we apply the suggested econometric methodology to the term structure of interest rates. We find strong evidence for threshold effects. Our model clearly outperforms the random walk model and the linear error correction model in terms of forecasting ability.

However, many issues remain for future research. Among others, working out a distribution theory for the parameter estimates is probably challenging due to the non-standard distribution of the threshold estimators, see e.g. Chan (1993) and Hansen (2000). Furthermore, estimation and testing methods for deterministic components is an important issue. Finally, we would need a test to decide between a two-regime model and a three-regime model.

Appendix

Preliminary Lemmata

Lemma 2.1:

Suppose Assumptions (A1)-(A4) hold. Then as $T \to \infty$

(a) $\frac{Z_\beta Z_\beta^\top}{T} \to_p E[\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] \equiv Q$,

(b) $\frac{Z_{1,\beta} Z_{1,\beta}^\top}{T} \to_p E[1(\phi_{t-1} \leq r)\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] \equiv Q_1(r)$,

(c) $\frac{Z_{3,\beta} Z_{3,\beta}^\top}{T} \to_p E[1(\phi_{t-1} > r)\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] \equiv Q_3(r)$,

(d) $\frac{Z_{2,\beta} Z_{2,\beta}^\top}{T} \to_p E[1(r_1 < \phi_{t-1} \leq r_2)\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] \equiv Q_2(r_1, r_2)$,

(e) $\tilde{\Omega} \to_p \Omega$,

where $Q, Q_1(r), Q_2(r_1, r_2)$ and $Q_3(r_2)$ are positive definite $(p \cdot k + r + 1) \times (p \cdot k + r + 1)$ matrices. Note that $Q = Q_1(1)$.

Proof:

Observe that all processes in ξ_t are stationary. Then, part (a) and (e) follow directly from the ergodic theorem. Part (b) and (c) follow from Theorem 3 in

Hansen (1996). Next, we show part (d). Using $Z_i Z_j^\top = 0$ $(i \neq j)$, we have $ZZ^\top = Z_1 Z_1^\top + Z_2 Z_2^\top + Z_3 Z_3^\top$. By part (a), (b) and (c), we have

$$\begin{aligned}
\frac{Z_{2,\beta} Z_{2,\beta}^\top}{T} &= \frac{Z_\beta Z_\beta^\top}{T} - \frac{Z_{1,\beta} Z_{1,\beta}^\top}{T} - \frac{Z_{3,\beta} Z_{3,\beta}^\top}{T} \\
&\rightarrow_p E[\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] - E[1(\phi_{t-1} \leq r_1)\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] - E[1(\phi_{t-1} > r_2)\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] \\
&= E[1(r_1 < \phi_{t-1} \leq r_2)\xi_{t-1,\beta}\xi_{t-1,\beta}^\top] \equiv Q_2(r_1, r_2).
\end{aligned}$$

by the Continuous Mapping Theorem.

Define

$$\begin{aligned}
H_{1,T}(r_1) &\equiv \frac{1}{\sqrt{T}}(Z_{1,\beta} \otimes I_p)\text{vec } U, \\
H_{2,T}(r_1, r_2) &\equiv \frac{1}{\sqrt{T}}(Z_{2,\beta} \otimes I_p)\text{vec } U, \\
H_{3,T}(r_2) &\equiv \frac{1}{\sqrt{T}}(Z_{3,\beta} \otimes I_p)\text{vec } U
\end{aligned}$$

Lemma 2.2:

Suppose Assumptions (A1)-(A4) hold. Then as $T \to \infty$

(a) $H_{1,T}(r_1) \Rightarrow H_1(r_1)$, where $H_1(r_1)$ is a $p(p \cdot k + r + 1)$-dimensional zero mean Gaussian process with covariance kernel $Q_1\left(r_1^{(1)} \wedge r_1^{(2)}\right) \otimes \Omega$,

(b) $H_{2,T}(r_1, r_2) \Rightarrow H_2(r_1, r_2)$, where $H_2(r_1, r_2)$ is a $p(p \cdot k + r + 1)$-dimensional zero mean Gaussian process with covariance kernel $Q_2\left(r_1^{(1)} \vee r_1^{(2)}, r_2^{(1)} \wedge r_2^{(2)}\right) \otimes \Omega$,

(c) $H_{3,T}(r_2) \Rightarrow H_3(r_2)$, where $H_3(r_2)$ is a $p(p \cdot k + r + 1)$-dimensional zero mean Gaussian process with covariance kernel $Q_3\left(r_2^{(1)} \vee r_2^{(2)}\right) \otimes \Omega$,

where $\vee$ and $\wedge$ denotes the max- and min-operator, respectively.

Proof:

Note that $(u_t \otimes \xi_{t-1} 1(\cdot))_{t=1,\ldots,T}$ is a $p(p \cdot k + r + 1)$-dimensional martingale difference sequence. By applying the central limit theorem for vector martingale difference sequences (e.g. see Hamilton, 1994, p. 194), we get the finite dimensional distributional convergence. It is well known that this is not sufficient to get weak convergence. Stochastic equicontinuity (see e.g. Andrews, 1994) is established by following the same steps as in Hansen (1996, Theorem 1). Stochastic equicontinuity, combined with the finite dimensional distributional convergence leads to weak convergence.

Next, we consider the case that y_t is nonstationary. We decompose $Z_l^*(r_{l-1}, r_l) = \left(Z_{1,l}^*(r_{l-1}, r_l), Z_{2,l}^*(r_{l-1}, r_l)\right)$ where

$$\begin{aligned} Z_{1,l}^* \equiv Z_{1,l}^*(r_{l-1}, r_l) &= \left(\xi_{1,0} 1(r_{l-1} < \phi_0 \leq r_l), \ldots, \xi_{1,T-1} 1(r_{l-1} < \phi_{T-1} \leq r_l)\right), \\ Z_{2,l}^* \equiv Z_{2,l}^*(r_{l-1}, r_l) &= \left(\xi_{2,0} 1(r_{l-1} < \phi_0 \leq r_l), \ldots, \xi_{2,T-1} 1(r_{l-1} < \phi_{T-1} \leq r_l)\right), \\ \xi_{1,t-1} &= (1, T^{-1/2} \Omega^{-1/2} C(1) y_{t-1}^\top)^\top, \\ \xi_{2,t-1} &= (\Delta y_{t-1}^\top, \ldots, \Delta y_{t-k}^\top)^\top. \end{aligned}$$

Note that under the null hypothesis y_t is generated by the stochastic process $\Delta y_t = C(L) u_t$ where $C(L) = C(1) + C_1(L)(1-L)$.

Lemma 2.3:

Under Assumptions (A1)-(A4), on $r_1, r_2 \in [0,1]$, $r_1 < r_2$,

(a) $\frac{Z_{2,1}^* Z_{2,1}^{*\top}}{T} \to_p E\left(1(\phi_{t-1} \leq r_1) \xi_{2,t-1} \xi_{2,t-1}^\top\right) \equiv Q_{22,1}(r_1)$,

(b) $\frac{Z_{2,3}^* Z_{2,3}^{*\top}}{T} \to_p E\left(1(\phi_{t-1} > r_2) \xi_{2,t-1} \xi_{2,t-1}^\top\right) \equiv Q_{22,3}(r_2)$,

(c) $\frac{Z_{2,2}^* Z_{2,2}^{*\top}}{T} \to_p E\left(1(r_1 < \phi_{t-1} \leq r_2) \xi_{2,t-1} \xi_{2,t-1}^\top\right) \equiv Q_{22,2}(r_1, r_2)$,

as $T \to \infty$.

Proof:

Note that Assumption 2 in Caner and Hansen (2001) is directly implied by Assumptions (A1)-(A4). Part (a) follows from the second part of Theorem 3 in Caner and

Hansen (2001). Next, we prove part (b). From the ergodic theorem and part (a) we get

$$\begin{aligned}
\frac{Z_{2,3}^{*}Z_{2,3}^{*\top}}{T} &= \frac{Z_{2}^{*}Z_{2}^{*\top}}{T} - \frac{Z_{2,1}^{*}(r_2)Z_{2,1}^{*\top}(r_2)}{T} \\
&\to_p E\left(\xi_{2,t-1}\xi_{2,t-1}^{\top}\right) - E\left(1(\phi_{t-1} \leq r_2)\xi_{2,t-1}\xi_{2,t-1}^{\top}\right) \\
&= E\left(1(\phi_{t-1} > r_2)\xi_{2,t-1}\xi_{2,t-1}^{\top}\right),
\end{aligned}$$

where $Z_2^{*} = (\xi_{2,0}, \ldots, \xi_{2,T-1})$. Finally, part (c) is shown in a similar way as part (b) by using the previous parts.

Lemma 2.4:

Let $X(s) = (1, W(s)^{\top})^{\top}$, where $W(s)$ is a p-dimensional Brownian motion. Then, under Assumptions (A1)-(A4), on $r_1, r_2 \in [0,1]$, $r_1 < r_2$,

(a) $\frac{Z_{1,1}^{*}Z_{2,1}^{*\top}}{T} \Rightarrow E\left(1(\phi_{t-1} \leq r_1)\xi_{2,t-1}\right)\int_0^1 X(s)^{\top}ds \equiv Q_{12,1}(r_1)$,

(b) $\frac{Z_{1,3}^{*}Z_{2,3}^{*\top}}{T} \Rightarrow E\left(1(\phi_{t-1} > r_2)\xi_{2,t-1}\right)\int_0^1 X(s)^{\top}ds \equiv Q_{12,3}(r_2)$,

(c) $\frac{Z_{1,2}^{*}Z_{2,2}^{*\top}}{T} \Rightarrow E\left(1(r_1 < \phi_{t-1} \leq r_2)\xi_{2,t-1}\right)\int_0^1 X(s)^{\top}ds$
$\equiv Q_{12,2}(r_1, r_2)$,

as $T \to \infty$.

Proof:

The first part of Theorem 3 in Caner and Hansen (2001) leads to

$$\lambda^{\top}\frac{Z_{1,1}^{*}Z_{2,1}^{*\top}}{T} \Rightarrow \lambda^{\top}E\left(1(\phi_{t-1} \leq r_1)\xi_{2,t-1}\right)\int_0^1 X(s)^{\top}ds,$$

where $\lambda \in \mathbb{R}^p$ and $\lambda^{\top}\lambda = 1$. Then, the result follows by the functional Cramér-Wold device, see e.g. Wooldridge and White (1988), Proposition 4.1. Next, we show part (b). Observe that

$$\begin{aligned}
\frac{Z_{1,3}^{*}Z_{2,3}^{*\top}}{T} &= \frac{Z_{1}^{*}Z_{2}^{*\top}}{T} - \frac{Z_{1,1}^{*}(r_2)Z_{2,1}^{*\top}(r_2)}{T} \\
&\Rightarrow E\left(\xi_{2,t-1}\right)\int_0^1 X(s)^\top ds - E\left(1(\phi_{t-1} \leq r_2)\xi_{2,t-1}\right)\int_0^1 X(s)^\top ds \\
&= E\left(1(\phi_{t-1} > r_2)\xi_{2,t-1}\right)\int_0^1 X(s)^\top ds,
\end{aligned}$$

with $Z_1^* := (\xi_{1,0}, \ldots, \xi_{1,T-1})$. Weak convergence follows from Phillips and Durlauf (1986) and part (a). Finally, part (c) is shown in a similar way as part (b) by using the previous parts.

Lemma 2.5:

Under Assumptions (A1)-(A4), on $r_1, r_2 \in [0,1]$, $r_1 < r_2$,

(a) $\frac{Z_{1,1}^{*}Z_{1,1}^{*\top}}{T} \Rightarrow r_1 \int_0^1 X(s)X(s)^\top ds \equiv Q_{11,1}(r_1)$,

(b) $\frac{Z_{1,3}^{*}Z_{1,3}^{*\top}}{T} \Rightarrow (1-r_2)\int_0^1 X(s)X(s)^\top ds \equiv Q_{11,3}(r_2)$,

(c) $\frac{Z_{1,2}^{*}Z_{1,2}^{*\top}}{T} \Rightarrow (r_2-r_1)\int_0^1 X(s)X(s)^\top ds \equiv Q_{11,2}(r_1,r_2)$,

as $T \to \infty$.

Proof:

The third part of Theorem 3 in Caner and Hansen (2001) leads to

$$\lambda^\top \frac{Z_{1,1}^{*}Z_{1,1}^{*\top}}{T} \Rightarrow r_1\lambda^\top \int_0^1 X(s)X(s)^\top ds,$$

where $\lambda \in \mathbb{R}^{p+1}$ and $\lambda^\top\lambda = 1$. Then, the result follows by the functional Cramér-Wold device, see e.g. Wooldridge and White (1988), Proposition 4.1. Next, we show part (b). Observe that

$$
\begin{aligned}
\frac{Z_{1,3}^{*}Z_{1,3}^{*\top}}{T} &= \frac{Z_{1}^{*}Z_{1}^{*\top}}{T} - \frac{Z_{1,1}^{*}(r_2)Z_{1,1}^{*\top}(r_2)}{T} \\
&\Rightarrow \int_0^1 X(s)X(s)^{\top}ds - r_2\int_0^1 X(s)X(s)^{\top}ds \\
&= (1-r_2)\int_0^1 X(s)X(s)^{\top}ds,
\end{aligned}
$$

where Z_1^* is defined as in the proof of Lemma 2.4. Weak convergence follows from Phillips and Durlauf (1986) and part (a). Finally, part (c) is shown in a similar way as part (b) by using the previous parts.

Applying Lemma 2.3-2.5 we obtain

$$
\begin{aligned}
Z_l^*(r_{l-1},r_l)Z_l^*(r_{l-1},r_l)^{\top} &= (Z_{1,l}^*(r_{l-1},r_l), Z_{2,l}^*(r_{l-1},r_l))\begin{pmatrix} Z_{1,l}^*(r_{l-1},r_l)^{\top} \\ Z_{2,l}^*(r_{l-1},r_l)^{\top} \end{pmatrix} \\
&= \begin{pmatrix} Z_{1,l}^*(r_{l-1},r_l)Z_{1,l}^*(r_{l-1},r_l)^{\top} & Z_{2,l}^*(r_{l-1},r_l)Z_{1,l}^*(r_{l-1},r_l)^{\top} \\ Z_{1,l}^*(r_{l-1},r_l)Z_{2,l}^*(r_{l-1},r_l)^{\top} & Z_{2,l}^*(r_{l-1},r_l)Z_{2,l}^*(r_{l-1},r_l)^{\top} \end{pmatrix} \\
&\Rightarrow \begin{pmatrix} Q_{11,l}(r_{l-1},r_l) & Q_{12,l}(r_{l-1},r_l) \\ Q_{12,l}(r_{l-1},r_l) & Q_{22,l}(r_{l-1},r_l) \end{pmatrix} \equiv Q_l^*(r_{l-1},r_l).
\end{aligned}
$$

with $l = 1, 2, 3$.

Proof of Theorem 2.1

We define the empirical residuals

$$\hat{U}(\gamma_1,\gamma_2) = \Delta Y - \hat{\Pi}_1 Z_1 - \hat{\Pi}_2 Z_2 - \hat{\Pi}_3 Z_3.$$

In view to (2.11) we have

$$\begin{aligned}
\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top &= (\Delta Y - \hat{\Pi}_1 Z_1 - \hat{\Pi}_2 Z_2 - \hat{\Pi}_3 Z_3)(\Delta Y^\top - Z_1^\top \hat{\Pi}_1^\top - Z_2^\top \hat{\Pi}_2^\top - Z_3^\top \hat{\Pi}_3^\top) \\
&= \Delta Y \Delta Y^\top - \Delta Y Z_1^\top \hat{\Pi}_1^\top - \Delta Y Z_2^\top \hat{\Pi}_2^\top - \Delta Y Z_3^\top \hat{\Pi}_3^\top \\
&- \hat{\Pi}_1 Z_1 \Delta Y^\top + \hat{\Pi}_1 Z_1 Z_1^\top \hat{\Pi}_1^\top + \hat{\Pi}_1 Z_1 Z_2^\top \hat{\Pi}_2^\top + \hat{\Pi}_1 Z_1 Z_3^\top \hat{\Pi}_3^\top \\
&- \hat{\Pi}_2 Z_2 \Delta Y^\top + \hat{\Pi}_2 Z_2 Z_1^\top \hat{\Pi}_1^\top + \hat{\Pi}_2 Z_2 Z_2^\top \hat{\Pi}_2^\top + \hat{\Pi}_2 Z_2 Z_3^\top \hat{\Pi}_3^\top \\
&- \hat{\Pi}_3 Z_3 \Delta Y^\top + \hat{\Pi}_3 Z_3 Z_1^\top \hat{\Pi}_1^\top + \hat{\Pi}_3 Z_3 Z_2^\top \hat{\Pi}_2^\top + \hat{\Pi}_3 Z_3 Z_3^\top \hat{\Pi}_3^\top
\end{aligned}$$

Using the orthogonality of Z_i and Z_j $(i \neq j)$, i.e. $Z_i Z_j^\top = 0$, $\hat{\Pi}_1 = \Delta Y Z_i^\top (Z_i Z_i^\top)^{-1}$ $(i = 1,2,3)$, (see (2.4)-(2.6)) and introducing the projection matrix $M_i := Z_i^\top (Z_i Z_i^\top)^{-1} Z_i$ $(i = 1,2,3)$, we get

$$\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top = \Delta Y \Delta Y^\top - \Delta Y M_1 \Delta Y^\top - \Delta Y M_2 \Delta Y^\top - \Delta Y M_3 \Delta Y^\top.$$

Next, we denote the true threshold parameters by γ_1^0 and γ_2^0 and write the model evaluated at γ_1^0 and γ_2^0 as $\Delta Y = \Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U$, with $Z_1^0 = (\xi_0 I(q_0 \leq \gamma_1^0), \ldots, \xi_{T-1} I(q_{T-1} \leq \gamma_1^0))$, $Z_2^0 = (\xi_0 I(\gamma_1^0 < q_0 \leq \gamma_2^0), \ldots, \xi_{T-1} I(\gamma_1^0 < q_{T-1} \leq \gamma_2^0))$ and $Z_3 = (\xi_0 I(q_0 > \gamma_2^0), \ldots, \xi_{T-1} I(q_{T-1} > \gamma_2^0))$. Hence, it follows

$$\begin{aligned}
\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top &= (\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U)(\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U)^\top \\
&- (\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U) M_1 (\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U)^\top \\
&- (\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U) M_2 (\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U)^\top \\
&- (\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U) M_3 (\Pi_1 Z_1^0 + \Pi_2 Z_2^0 + \Pi_3 Z_3^0 + U)^\top.
\end{aligned}$$

It remains to show that the objective function converges uniformly in probability to a nonstochastic limit that is uniquely minimized at $(\gamma_1, \gamma_2) =: \gamma = \gamma^0 := (\gamma_1^0, \gamma_2^0)$. Under the restrictions $\gamma_1 < \gamma_2$ and $\gamma_1^0 < \gamma_2^0$, it is sufficient to consider the following six cases. Consider first the case that y_t is stationary. Applying appropriate normalizations, we obtain by Lemma 2.1 after some lengthy calculations

Case (1): $\gamma_1 < \gamma_1^0 < \gamma_2 < \gamma_2^0$

$$\begin{aligned}\frac{\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top}{T} &\to_p (\Pi_1-\Pi_2)Q_2(\gamma_1,\gamma_1^0)Q_2^{-1}(\gamma_1,\gamma_2)Q_2(\gamma_1^0,\gamma_2)(\Pi_1-\Pi_2)^\top \\ &+ (\Pi_2-\Pi_3)Q_2(\gamma_2,\gamma_2^0)Q_3^{-1}(\gamma_2)Q_3(\gamma_2^0)(\Pi_2-\Pi_3)^\top + \Omega \\ &> \Omega,\end{aligned}$$

Case (2): $\gamma_1 < \gamma_1^0 < \gamma_2^0 < \gamma_2$

$$\begin{aligned}\frac{\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top}{T} &\to_p (\Pi_1-\Pi_2)Q_2(\gamma_1,\gamma_1^0)Q_2^{-1}(\gamma_1,\gamma_2)Q_2(\gamma_1^0,\gamma_2^0)(\Pi_1-\Pi_2)^\top \\ &+ (\Pi_1-\Pi_3)Q_2(\gamma_1,\gamma_1^0)Q_2^{-1}(\gamma_1,\gamma_2)Q_2(\gamma_2^0,\gamma_2)(\Pi_1-\Pi_3)^\top \\ &+ (\Pi_2-\Pi_3)Q_2(\gamma_1^0,\gamma_2^0)Q_2^{-1}(\gamma_1,\gamma_2)Q_2(\gamma_2^0,\gamma_2)(\Pi_2-\Pi_3)^\top + \Omega \\ &> \Omega,\end{aligned}$$

Case (3): $\gamma_1^0 < \gamma_1 < \gamma_2 < \gamma_2^0$

$$\begin{aligned}\frac{\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top}{T} &\to_p (\Pi_1-\Pi_2)Q_1(\gamma_1^0)Q_1^{-1}(\gamma_1)Q_2(\gamma_1^0,\gamma_1)(\Pi_1-\Pi_2)^\top \\ &+ (\Pi_2-\Pi_3)Q_2(\gamma_2,\gamma_2^0)Q_3^{-1}(\gamma_2)Q_3(\gamma_2^0)(\Pi_2-\Pi_3)^\top + \Omega \\ &> \Omega,\end{aligned}$$

Case (4): $\gamma_1^0 < \gamma_1 < \gamma_2^0 < \gamma_2$

$$
\begin{aligned}
\frac{\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top}{T} &\to_p (\Pi_1-\Pi_2)Q_1(\gamma_1^0)Q_1^{-1}(\gamma_1)Q_2(\gamma_1^0,\gamma_1)(\Pi_1-\Pi_2)^\top \\
&+ (\Pi_2-\Pi_3)Q_2(\gamma_1,\gamma_2^0)Q_2^{-1}(\gamma_1,\gamma_2)Q_2(\gamma_2^0,\gamma_2)(\Pi_2-\Pi_3)^\top + \Omega \\
&> \Omega,
\end{aligned}
$$

Case (5): $\gamma_1 < \gamma_2 < \gamma_1^0 < \gamma_2^0$

$$
\begin{aligned}
\frac{\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top}{T} &\to_p (\Pi_1-\Pi_2)Q_2(\gamma_2,\gamma_1^0)Q_3^{-1}(\gamma_2)Q_2(\gamma_1^0,\gamma_2^0)(\Pi_1-\Pi_2)^\top \\
&+ (\Pi_2-\Pi_3)Q_2(\gamma_1^0,\gamma_2^0)Q_3^{-1}(\gamma_2)Q_3(\gamma_2^0)(\Pi_2-\Pi_3)^\top + \Omega \\
&> \Omega,
\end{aligned}
$$

Case (6): $\gamma_1^0 < \gamma_2^0 < \gamma_1 < \gamma_2$

$$
\begin{aligned}
\frac{\hat{U}(\gamma_1,\gamma_2)\hat{U}(\gamma_1,\gamma_2)^\top}{T} &\to_p (\Pi_1-\Pi_2)Q_1(\gamma_1^0)Q_1^{-1}(\gamma_1)Q_2(\gamma_1^0,\gamma_2^0)(\Pi_1-\Pi_2)^\top \\
&+ (\Pi_1-\Pi_3)Q_1(\gamma_1^0)Q_1^{-1}(\gamma_1)Q_2(\gamma_2^0,\gamma_1)(\Pi_1-\Pi_3)^\top \\
&+ (\Pi_2-\Pi_3)Q_2(\gamma_1^0,\gamma_2^0)Q_1^{-1}(\gamma_1)Q_2(\gamma_2^0,\gamma_1)(\Pi_2-\Pi_3)^\top + \Omega \\
&> \Omega,
\end{aligned}
$$

If y_t is nonstationary, the results continue to be valid by replacing $Q_l(\gamma_{l-1},\gamma_l)$ by $Q_l^*(\gamma_{l-1},\gamma_l)$.

An alternative representation of the supLM statistic

First, we derive an alternative algebraic representation of $\mathrm{vec}(\hat{\delta}_1^\top, \hat{\delta}_3^\top)^\top$ being an essential building block of the supLM statistic. This representation will enable us to apply Lemma 2.1 and Lemma 2.2 to prove Theorem 2.3.

Let $\tilde{U} = \Delta Y - \hat{A}_\alpha Z_{\hat{\beta}} = \Delta Y - \Delta Y Z_{\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} Z_{\hat{\beta}}$ the residuals under H_0. Then, we have

$$
\begin{aligned}
\mathrm{vec}\ \hat{\delta}_{1,\alpha} &= \mathrm{vec}\ (\hat{A}_{1,\alpha} - \hat{A}_{2,\alpha}) \\
&= [(Z_{1,\hat{\beta}} Z_{1,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{1,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \Delta Y - [(Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{2,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \Delta Y \\
&= [(Z_{1,\hat{\beta}} Z_{1,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{1,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \tilde{U} + [(Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{1,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \tilde{U} \\
&+ [(Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{3,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \tilde{U}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{vec}\ \hat{\delta}_{3,\alpha} &= \mathrm{vec}\ (\hat{A}_{3,\alpha} - \hat{A}_{2,\alpha}) \\
&= [(Z_{3,\hat{\beta}} Z_{3,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{3,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \Delta Y - [(Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{2,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \Delta Y \\
&= [(Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{1,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \tilde{U} + [(Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{3,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \tilde{U} \\
&+ [(Z_{3,\hat{\beta}} Z_{3,\hat{\beta}}^\top)^{-1} \otimes I_p][Z_{3,\hat{\beta}} \otimes I_p]\mathrm{vec}\ \tilde{U}.
\end{aligned}
$$

This can be written as

$$
\begin{aligned}
\mathrm{vec}\begin{pmatrix} \hat{A}_{1,\alpha} - \hat{A}_{2,\alpha} \\ \hat{A}_{3,\alpha} - \hat{A}_{2,\alpha} \end{pmatrix} &= \begin{pmatrix} [(Z_{1,\hat{\beta}} Z_{1,\hat{\beta}}^\top)^{-1} + (Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1}] & (Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} \\ (Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} & [(Z_{2,\hat{\beta}} Z_{2,\hat{\beta}}^\top)^{-1} + (Z_{3,\hat{\beta}} Z_{3,\hat{\beta}}^\top)^{-1}] \end{pmatrix} \otimes I_p \\
&\cdot \begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \mathrm{vec}\ \tilde{U} \\
&= (M_T(r_1, r_2) \otimes I_p) \begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \mathrm{vec}\ \tilde{U}. \qquad (2.17)
\end{aligned}
$$

Using $\tilde{U} = \Delta Y - \hat{A}_\alpha Z_{\hat{\beta}} = \Delta Y - \Delta Y Z_{\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} Z_{\hat{\beta}}$, we get for the second term in (2.17)

$$
\begin{aligned}
\begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \text{vec } \tilde{U} &= \begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \text{vec } (\Delta Y - \Delta Y Z_{\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} Z_{\hat{\beta}}) \\
&= \begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \text{vec } \Delta Y \\
&- \begin{pmatrix} Z_{1,\hat{\beta}} Z_{1,\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} \otimes I_p \\ Z_{3,\hat{\beta}} Z_{3,\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} \otimes I_p \end{pmatrix} (Z_{\hat{\beta}} \otimes I_p) \text{vec } \Delta Y.
\end{aligned}
$$

For $l = 1, 2, 3$ we note that

$$
I(\gamma_{l-1} < q_{t-1} \leq \gamma_l) = I(r_{l-1} < \phi_{t-1} \leq r_l)
$$

with $\phi_{t-1} := F(q_{t-1})$. Then, we replace $I(\gamma_{l-1} < q_{t-1} \leq \gamma_l)$ by $I(r_{l-1} < \phi_{t-1} \leq r_l)$ in the definition of $Z_{l,\hat{\beta}}$ and define

$$
\text{supLM} = \sup_{\theta \leq r_1, r_2 \leq 1-\theta} \text{LM}(r_1, r_2).
$$

Under H_0, it holds that $\Delta Y = A_\alpha Z_\beta + U$. Hence, we obtain,

$$
\begin{aligned}
\begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \text{vec } \tilde{U} &= \begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \text{vec } (A_\alpha Z_\beta + U) \\
&- \begin{pmatrix} Z_{1,\hat{\beta}} Z_{1,\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} \otimes I_p \\ Z_{3,\hat{\beta}} Z_{3,\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} \otimes I_p \end{pmatrix} (Z_{\hat{\beta}} \otimes I_p) \text{vec } (A_\alpha Z_\beta + U) \\
&= \begin{pmatrix} Z_{1,\hat{\beta}} \otimes I_p \\ Z_{3,\hat{\beta}} \otimes I_p \end{pmatrix} \text{vec } U \\
&- \begin{pmatrix} Z_{1,\hat{\beta}} Z_{1,\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} \otimes I_p \\ Z_{3,\hat{\beta}} Z_{3,\hat{\beta}}^\top (Z_{\hat{\beta}} Z_{\hat{\beta}}^\top)^{-1} \otimes I_p \end{pmatrix} (Z_{\hat{\beta}} \otimes I_p) \text{vec } U \\
&=: J_T(r_1, r_2). \qquad (2.18)
\end{aligned}
$$

Hence, we have by (2.17) and (2.18)

$$\operatorname{vec}\begin{pmatrix} \hat{A}_{1,\alpha} - \hat{A}_{2,\alpha} \\ \hat{A}_{3,\alpha} - \hat{A}_{2,\alpha} \end{pmatrix} = (M_T(r_1, r_2) \otimes I_p) J_T(r_1, r_2).$$

Proof of Theorem 2.3

Since $(\hat{\beta} - \beta) = O_p(T^{-1})$ observe that $|\xi_{t,\hat{\beta}} - \xi_{t,\beta}| = |y_t^\top(\hat{\beta} - \beta)| = O_p(T^{-1/2})$, where $|\cdot|$ denotes an arbitrary norm. Therefore, we can replace in the remainder of the proof $Z_{\hat{\beta}}$ by Z_β. Consider the matrix $M_T(r_1, r_2)$. By applying Lemma 2.1, we get

$$
\begin{aligned}
T \cdot M_T(r_1, r_2) = T \cdot &\begin{pmatrix} (Z_{1,\beta} Z_{1,\beta}^\top)^{-1} + (Z_{2,\beta} Z_{2,\beta}^\top)^{-1} & (Z_{2,\beta} Z_{2,\beta}^\top)^{-1} \\ (Z_{2,\beta} Z_{2,\beta}^\top)^{-1} & (Z_{2,\beta} Z_{2,\beta}^\top)^{-1} + (Z_{3,\beta} Z_{3,\beta}^\top)^{-1} \end{pmatrix} \\
= \; &\begin{pmatrix} (T^{-1} \cdot Z_{1,\beta} Z_{1,\beta}^\top)^{-1} + (T^{-1} \cdot Z_{2,\beta} Z_{2,\beta}^\top)^{-1} & (T^{-1} \cdot Z_{2,\beta} Z_{2,\beta}^\top)^{-1} \\ (T^{-1} \cdot Z_{2,\beta} Z_{2,\beta}^\top)^{-1} & (T^{-1} \cdot Z_{2,\beta} Z_2^\top)^{-1} + (T^{-1} \cdot Z_{3,\beta} Z_{3,\beta}^\top)^{-1} \end{pmatrix} \\
\to_p \; &\begin{pmatrix} Q_1(r_1)^{-1} + Q_2(r_1, r_2)^{-1} & Q_2(r_1, r_2)^{-1} \\ Q_2(r_1, r_2)^{-1} & Q_2(r_1, r_2)^{-1} + Q_3(r_2)^{-1} \end{pmatrix} \\
=: \; & M(r_1, r_2)
\end{aligned}
\tag{2.19}
$$

Next, consider the term $J_T(r_1, r_2)$. By introducing the $\sqrt{T}$-normalization, we obtain the distributional result

$$
\begin{aligned}
\frac{1}{\sqrt{T}} J_T(r_1, r_2) &= \frac{1}{\sqrt{T}} \begin{pmatrix} Z_{1,\beta} \otimes I_p \\ Z_{3,\beta} \otimes I_p \end{pmatrix} \operatorname{vec} \tilde{U} = \frac{1}{\sqrt{T}} \begin{pmatrix} Z_{1,\beta} \otimes I_p \\ Z_{3,\beta} \otimes I_p \end{pmatrix} \operatorname{vec} U \\
&- \begin{pmatrix} T^{-1} Z_{1,\beta} Z_{1,\beta}^\top (T^{-1} Z_\beta Z_\beta^\top)^{-1} \otimes I_p \\ T^{-1} Z_{3,\beta} Z_{3,\beta}^\top (T^{-1} Z_\beta Z_\beta^\top)^{-1} \otimes I_p \end{pmatrix} \frac{1}{\sqrt{T}} (Z_\beta \otimes I_p) \operatorname{vec} U \\
&\Rightarrow \begin{pmatrix} H_1(r_1) \\ H_3(r_2) \end{pmatrix} \\
&- \begin{pmatrix} Q_1(r_1) Q^{-1} \otimes I_p \\ Q_3(r_2) Q^{-1} \otimes I_p \end{pmatrix} H_1(1) \equiv J(r_1, r_2)
\end{aligned}
\tag{2.20}
$$

Finally, note that by (2.19)

$$\begin{aligned} T \cdot M_T(r_1, r_2) \otimes \tilde{\Omega} &\to_p \begin{pmatrix} Q_1(r_1)^{-1} + Q_2(r_1, r_2)^{-1} & Q_2(r_1, r_2)^{-1} \\ Q_2(r_1, r_2)^{-1} & Q_2(r_1, r_2)^{-1} + Q_3(r_2)^{-1} \end{pmatrix} \otimes \Omega \\ &\equiv M(r_1, r_2) \otimes \Omega \end{aligned}$$

by using Lemma 2.1, part (e) and (2.19). It follows by the continuous mapping theorem that

$$(T \cdot M_T(r_1, r_2))^{-1} \otimes \tilde{\Omega}^{-1} \to_p M(r_1, r_2)^{-1} \otimes \Omega^{-1}. \tag{2.21}$$

Combining (2.19), (2.20) and (2.21) leads to

$$\begin{aligned} & J(r_1, r_2)^\top (M(r_1, r_2) \otimes I_p)(M(r_1, r_2)^{-1} \otimes \Omega^{-1})(M(r_1, r_2) \otimes I_p) J(r_1, r_2) \\ = \; & J(r_1, r_2)^\top \left[M(r_1, r_2) \otimes \Omega^{-1} \right] J(r_1, r_2) \end{aligned}$$

as claimed in Theorem 2.3.

Chapter 3

A Partially Linear Approach to Modelling the Dynamics of Spot and Futures Prices

3.1 Introduction

Prices in spot and futures markets are linked through the cost-of-carry relation. In a frictionless world arbitrage would eliminate any deviations from this relation. In practice, however, such deviations may and do occur for several reasons. First, the existence of transactions costs makes it unprofitable to exploit small deviations. Second, traders with access to private information may prefer to trade in a specific market. Consequently, prices in this market may reflect information earlier than prices in the other market. As transaction costs tend to be lower in the futures market (e.g. Berkmann et al. 2005) informed traders may prefer to trade in this market and it thus might reflect the information earlier than the spot market. The opposite may also occur, however. Consider a trader with information on the value of an individual stock. The trader can trade on that information in the spot market. In the futures market, on the other hand, he is restricted to trading a basket of securities (i.e., an index futures contract). Therefore, firm-specific information may be reflected in the spot market first.

The question of which market impounds new information faster is thus an empirical one, and it has been subject to academic research for about two decades.[1] The empirical methods have been considerably refined since the early work of Kawaller et al. (1987) and others. VAR models were introduced (e.g. Stoll and Whaley 1990) and soon thereafter replaced by error correction (ECM) models. A standard ECM implicitly assumes that deviations of prices from their long-run equilibrium (the pricing errors) are reduced at a speed that is independent of the magnitude of the price deviation. This is unlikely to be the case, however. Whenever the deviations are sufficiently large to allow for profitable arbitrage, the speed of adjustment should increase.[2] Some authors (e.g. Yadav et al. 1994, Dwyer et al. 1996 and Martens et al. 1998) have employed threshold error correction (TECM) models to address this issue. A TECM assumes a non-continuous transition function and allows for a discrete number of different speed of adjustment coefficients. If all traders would face identical transaction costs, a TECM with two different adjustment coefficients (i.e., a no-arbitrage regime and an arbitrage regime) would be a reasonable choice. If, on the other hand, traders are heterogeneous with respect to the transaction costs they face, a less restrictive model is warranted. An obvious candidate is a smooth transition error correction (STECM) model as applied by Taylor et al. (2000), Anderson and Vahid (2001) and Tse (2001).

A shortcoming of the STECM models is that the transition function must be exogenously specified, and there is no theory to guide the specification of the model. The researcher also has to decide for a symmetric transition function or one that allows for asymmetry. Such asymmetries may arise because short sales in the spot market are more expensive than short sales in the futures market.

The contribution of this chapter is to propose a more flexible modelling frame-

[1]Given the nature of our empirical analysis we restrict the brief survey of the literature to papers analyzing the relation between stock price indices and stock index futures contracts.

[2]The width of the arbitrage bounds is likely to depend on the liquidity of the market. In a recent paper Roll et al. (2007) have documented a relation between liquidity and the futures-cash basis for the NYSE composite index futures contract over the period 1988-2002.

work. We estimate a partially linear ECM where the adjustment process is modelled non-parametrically. The short-run dynamics are estimated by density-weighted OLS based on the approach proposed by Fan and Li (1999a). The non-parametric function modelling the adjustment process is estimated by a Nadaraya-Watson estimator. The modelling approach that we use was proposed by Gaul (2005) but has as yet not been applied.

We implement our model using data from the German stock market. Specifically, we analyze the dynamics of the DAX index and the DAX futures contract. The results suggest that the speed of adjustment is indeed monotonically increasing in the magnitude of the price deviation. We test our specification against a standard ECM and clearly reject the latter. Estimates of the parameters governing the short-run dynamics are similar in the standard ECM and in our model.

These results have several implications. First, they confirm the intuition that the speed of adjustments of prices to deviations from equilibrium is increasing in the magnitude of the deviation. Second, they imply that a standard ECM as well as a TECM is unable to fully capture the dynamics of the adjustment process. Third, the form of the non-parametric adjustment function may guide the choice for a functional form in STECM models.

The remainder of the chapter is organized as follows. Section 3.2 provides a description of the data set. In section 3.3 we describe the estimation procedure. In section 3.4 we describe a test for linearity. Section 3.5 is devoted to the presentation of the results, section 3.6 concludes.

3.2 Market Structure and Data

Our analysis uses DAX index level data and bid and ask quotes from the DAX index futures contract traded on Eurex. The DAX is a value-weighted index calculated from the prices of the 30 largest German stocks. The prices are taken from Xetra, the most liquid market for German stocks.[3] Index values are published in intervals of 15 seconds. The DAX is a performance index, i.e., the calculation of the index is based on the presumption that dividends are reinvested. As a consequence, the expected dividend yield does not enter the cost of carry relation. Besides an index calculated from the most recent transaction prices the exchange also calculates an index from the current best ask prices (ADAX) and an index calculated from the current best bid prices (BDAX). These indices are value-weighted averages of the inside quotes, and their mean is equivalent to a value-weighted average of the quote midpoints of the component stocks.

Futures contracts on the DAX are traded on the EUREX. The contracts are cash-settled and trade on a quarterly cycle. They mature on the third Friday of the months March, June, September, and December. The DAX futures contract is a highly liquid instrument. In the first quarter of 1999 (our sample period), more than 1,150,000 transactions were recorded. The open interest at the end of the quarter was more than 290,000 contracts.

Both Xetra and EUREX are electronic open limit order books. Therefore, the results of our empirical analysis are unlikely to be affected by differences in market structure. The trading hours in the two markets are different, though. Trading in Xetra starts with a call auction held between 8.25 am and 8:30 am. After the opening auction, continuous trading starts and extends until 5 pm, interrupted by an intraday auction which takes place between 1:00 pm and 1:02 pm. Trading of the DAX futures contract starts at 9 am and extends until 5 pm.

We obtained all data from Bloomberg. Our sample period is the first quarter of 1999 and extends over 61 trading days. For this period we obtained the values of the DAX index and the two quote-based indices ADAX and BDAX at a frequency of 15 seconds. From the quote-based indices we calculate the midquote index $MQDAX_t = \frac{ADAX_t + BDAX_t}{2}$. We further obtained a time series of all bid and ask quotes and all transaction prices of the nearby DAX futures contract. We only use data for the period

[3]The DAX stocks are traded on Xetra, on the floor of the Frankfurt Stock Exchange and on several regional exchanges. The market share of Xetra amounted to 90% during our sample period.

of simultaneous operation of both markets. We further discard all observations before 9 am and from 4:55 pm onwards. We also discard all observations within 5 minutes from the time of the intraday call auction (held between 1:00 pm and 1:02 pm). After these adjustments the sample consists of 100188 observations.

All estimations are based on quote midpoints. They are preferred to transaction prices because the use of midpoints alleviates the infrequent trading problem.[4] We match each index level observation whith the bid and ask quotes in the futures market that were in effect at the time the index level information was published.

The cost-of-carry relation implies that the cash index and the futures contract are cointegrated. In order to eliminate the time-variation of the cointegrating relation we discount the futures prices using daily observations on the one-month interbank rate as published by Deutsche Bundesbank.[5]

As a prerequisite for our empirical analysis we have to establish that the time series are I(1) and are cointegrated. Table 3.1 presents the results of augmented Dickey-Fuller tests and Phillips-Perron tests applied to p_t and Δp_t. p_t denotes a log price series observed at date t and the indices X and F identify observations relating to the cash market (X, Xetra) and the futures market (F), respectively. Δ is the difference operator. The results of the stationarity tests clearly suggest that all series are I(1). In equilibrium spot and futures prices are linked through the cost-of-carry relation. Consequently, the DAX index level and the discounted futures price should be equal in equilibrium, and their difference should be stationary. We test the latter hypothesis using both an augmented Dickey-Fuller test and a Phillips-Perron test and clearly reject the null of a unit root (p-value 0.0000 and 0.0001, respectively). This result

[4]Spot market index levels are calculated using the last available transaction price for each of the component stocks. As stocks do not trade simultaneously, some of the prices used to calculate the index are stale. This may induce positive serial correlation in the index returns. Quote midpoints, on the other hand, are based on tradable bid and ask prices and should be less affected by the infrequent trading problem. See Shyy et al. (1996) or Theissen (2005).

[5]Given the margin requirements in the futures market, the rate for overnight deposits is an alternative choice. However, the time series of overnight deposit rates exhibits peaks which may be due to bank reserve requirements. Besides, the term structure at the short end was essentially flat during the sample period, making the choice of the interest rate less important.

Table 3.1: Results of the unit root tests for XDAX and FDAX

	Level		First Difference	
	Augmented DF	Phillips / Perron	Augmented DF	Phillips / Perron
p^X	0.5773	0.6395	0.0001	0.0001
p^F	0.3964	0.4113	0.0001	0.0001

confirms the theoretical prediciton that spot and futures prices are cointegrated with the cointegrating vector being $(1, -1)^\top$. We use this pre-specified cointegrating vector in our estimation.

3.3 Estimation procedure

For the reasons exposed in the Introduction, our model is characterized by a nonparametric function for the pricing error. In particular, we propose to use the model

$$\Delta y_t = \sum_{i=1}^{k} \Gamma_i \Delta y_{t-i} + F(\beta^\top y_{t-1}) + \epsilon_t, \qquad \text{t=1,...,T}, \tag{3.1}$$

where y_t denotes a vector process containing the variables p_t^X and p_t^F. The cointegrating vector is denoted by β and is pre-specified to $(1, -1)^\top$. The adjustment process is described by the unknown nonparametric function F: $\mathbb{R} \to \mathbb{R}^2$ and ϵ_t is a two-dimensional error process. By introducing the $2 \times 2k$-matrix $\Gamma := (\Gamma_1 \dots \Gamma_k)$ and the $2k$-dimensional vector $\xi_{t-1} := \left(\Delta y_{t-1}^\top \dots \Delta y_{t-k}^\top\right)^\top$, model (3.1) can be written as

$$\Delta y_t = \Gamma \xi_{t-1} + F(\beta^\top y_{t-1}) + \epsilon_t. \tag{3.2}$$

Note that model (3.2) contains the linear VECM (Engle and Granger, 1987; Johansen, 1988), the threshold VECM (Hansen and Seo, 2002) and the smooth transition VECM (van Dijk and Franses, 2000) as special cases.

The estimation procedure described in the following involves two stages. First, we estimate the matrix Γ, then the function F.

3.3.1 Estimation of Γ

Taking expectations in (3.2) conditional on $\beta^\top y_{t-1}$, we have

$$E(\Delta y_t|\beta^\top y_{t-1}) = \Gamma E(\xi_{t-1}|\beta^\top y_{t-1}) + F(\beta^\top y_{t-1}), \tag{3.3}$$

using $E(\epsilon_t|\beta^\top y_{t-1}) = 0$. Subtracting (3.3) from (3.2) leads to

$$\Delta y_t - E(\Delta y_t|\beta^\top y_{t-1}) = \Gamma(\xi_{t-1} - E(\xi_{t-1}|\beta^\top y_{t-1})) + \epsilon_t, \tag{3.4}$$

which has the following form

$$\Delta y_t^* = \Gamma \xi_{t-1}^* + \epsilon_t, \tag{3.5}$$

where $\Delta y_t^* := \Delta y_t - E(\Delta y_t|\beta^\top y_{t-1})$ and $\xi_{t-1}^* := \xi_{t-1} - E(\xi_{t-1}|\beta^\top y_{t-1})$. If $E(\Delta y_t|\beta^\top y_{t-1})$ and $E(\xi_{t-1}|\beta^\top y_{t-1})$ were known, Γ could be estimated by OLS. Since $E(\Delta y_t|\beta^\top y_{t-1})$ and $E(\xi_{t-1}|\beta^\top y_{t-1})$ are usually unknown, an estimator based on Δy_t^* and ξ_{t-1}^* is not feasible. To obtain a feasible estimator, we will use the nonparametric kernel method, similar to Robinson (1988) and Fan and Li (1999a). In particular, the conditional means $E(\Delta y_t|\beta^\top y_{t-1})$ and $E(\xi_{t-1}|\beta^\top y_{t-1})$ are estimated by the Nadaraya-Watson estimator

$$\begin{aligned}
\hat{E}(\Delta y_t|\beta^\top y_{t-1}) &= \frac{1}{Th}\sum_{j=1}^{T} \Delta y_j K\left(\frac{\beta^\top y_{t-1} - \beta^\top y_{j-1}}{h}\right)/\hat{f}(\beta^\top y_{t-1}),\\
\hat{E}(\xi_{t-1}|\beta^\top y_{t-1}) &= \frac{1}{Th}\sum_{j=1}^{T} \xi_{j-1} K\left(\frac{\beta^\top y_{t-1} - \beta^\top y_{j-1}}{h}\right)/\hat{f}(\beta^\top y_{t-1}),
\end{aligned}$$

where

$$\hat{f}(\beta^\top y_{t-1}) = \frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{\beta^\top y_{t-1} - \beta^\top y_{j-1}}{h}\right) \tag{3.6}$$

is the kernel density estimator for $f(\beta^\top y_{t-1})$, $K(\cdot)$ is a kernel function and h is a bandwidth parameter.

To avoid the random denominator problem in kernel estimation (i.e. the occurrence of small values of the estimated density function), we use density weighted estimates, similar to Fan and Li (1999a). Thus, we multiply (3.5) by $f(\beta^\top y_{t-1})$, the density function of $\beta^\top y_{t-1}$, and obtain

$$f(\beta^\top y_{t-1})\Delta y_t^* = \Gamma f(\beta^\top y_{t-1})\xi_{t-1}^* + f(\beta^\top y_{t-1})\epsilon_t. \tag{3.7}$$

We replace $E(\Delta y_t|\beta^\top y_{t-1})$, $E(\xi_{t-1}|\beta^\top y_{t-1})$ and $f(\beta^\top y_{t-1})$ in (3.7) by their estimates. This leads to the feasible estimator

$$\hat{\Gamma}^{\text{OLS}} = \left[\sum_{t=1}^{T} \Delta\hat{y}_t^* \hat{\xi}_{t-1}^{*\top} \hat{f}(\beta^\top y_{t-1})^2\right] \left[\sum_{t=1}^{T} \hat{\xi}_{t-1}^* \hat{\xi}_{t-1}^{*\top} f(\beta^\top y_{t-1})^2\right]^{-1}, \tag{3.8}$$

with $\Delta\hat{y}_t^* := \Delta y_t - \hat{E}(\Delta y_t|\beta^\top y_{t-1})$ and $\hat{\xi}_{t-1}^* := \xi_{t-1} - \hat{E}(\xi_{t-1}|\beta^\top y_{t-1})$. Besides some technical assumptions, we assume that $(\Delta y_t, \beta^\top y_{t-1})$ is β-mixing, $Th^2 \to \infty$ and $Th^8 \to 0$ for $T \to \infty$. Similar to Fan and Li (1999a), it can be shown that vec $(\hat{\Gamma}^{\text{OLS}} - \Gamma)$ is $\sqrt{T}$ consistent and asymptotically normally distributed.

3.3.2 Estimation of F

Substituting $\hat{\Gamma}^{\text{OLS}}$ for Γ in model (3.2), one obtains the nonparametric model

$$\Delta\tilde{y}_t = F(\beta^\top y_{t-1}) + u_t, \tag{3.9}$$

where $\Delta\tilde{y}_t := \Delta y_t - \hat{\Gamma}^{\text{OLS}}\xi_{t-1}$. Applying the Nadaraya-Watson estimator to (3.9), i.e.

$$\hat{F}(z) = \frac{\sum_{t=1}^{T} \Delta\tilde{y}_t K\left(\frac{z-\beta^\top y_{t-1}}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{z-\beta^\top y_{t-1}}{h}\right)}, \tag{3.10}$$

we get an estimator for the function F. It is well known that $\hat{F}(\cdot)$ has the same asymptotic distribution as if Γ were known. Later, we will use this statement for constructing pointwise confidence intervals.

3.3.3 Bandwidth Selection

In empirical applications we have to choose both the kernel function and the bandwidth parameter h. Whereas the influence of the kernel function is negligible, the choice of the bandwidth parameter plays a crucial role. Due to our enormous sample size, standard bandwidth selection procedures like cross-validation, are no longer applicable as the computational time increases at quadratic rate with the number of observations. In order to determine the bandwidth parameter h we use the method of *Weighted Averaging of Rounded Points (WARPing)* developed by Härdle and Scott (1992). This technique is based on discretizing the data first into a finite grid of bins, then smoothing the binned data and finally selecting the optimal bandwidth using the binned data. The main advantage of WARPing is the substantial gain of computational efficiency. In particular, Härdle (1991) and Härdle and Scott (1992) show that the number of iterations increases at linear rate with the number of observations rather than quadratic.

In our application we determine the optimal bandwidth by using four different criteria, namely cross-validation, the Shibata's Model Selector, Akaike's Information Criterion and Final Prediction Error Criterion. For a detailed discussion of them, we refer to Härdle, Müller, Sperlich and Werwatz (2004). The lower limit for h for the grid search is set to 0.000332, the upper to 0.005307 and the bindwidth d to $6.634 \cdot 10^{-5}$. The number of equidistant grid points is chosen to be 100. The analysis is carried out by using the software package XploRe. The results are given in the following table.

Table 3.2: Results of bandwidth selection

Bandwidth selection procedure	XDAX	FDAX
Cross Validation	0.000371	0.000492
Shibata's Model Selector	0.000351	0.000492
Akaike's Information Criterion	0.000361	0.000492
Final Prediction Error	0.000361	0.000492

The table shows that all methods lead to very similar results for the XDAX series. According to Akaike's Information Criterion and Final Prediction Error we choose $h^X = 0.000361$. For the FDAX series, all methods yield the same result. Hence, we choose $h^F = 0.000492$.

3.4 Test for linearity

The linear vector error correction model

$$\Delta y_t = \Gamma \xi_{t-1} + \alpha \beta^\top y_{t-1} + \epsilon_t \tag{3.11}$$

may be considered the baseline model in cointegration analysis. We now provide a statistical single-equation test to examine the hypothesis whether model (3.11) is as accurate a description of the data as model (3.1). Formally, we are interested in testing the hypotheses

H_0 : $E(\Delta y_{it} | \xi_{t-1}, \beta^\top y_{t-1}) = \Gamma_i \xi_{t-1} + \alpha_i \beta^\top y_{t-1}$ for some Γ_i and α_i against

H_1 : $E(\Delta y_{it} | \xi_{t-1}, \beta^\top y_{t-1}) = \Gamma_i \xi_{t-1} + F_i(\beta^\top y_{t-1})$ with $P(F_i(\beta^\top y_{t-1}) = \alpha_i \beta^\top y_{t-1}) < 1)$ for any $\alpha_i \in \mathbb{R}$.

To motivate an appropriate test statistic, we consider (3.2) with $\Gamma = 0$. Denote $u_{it} := \Delta y_{it} - \alpha_i \beta^\top y_{t-1}$ the residuals under H_0. Following Zheng (1996) and Li and Wang (1998), our test is based on $E\left[u_{it} E[u_{it} | \beta^\top y_{t-1}] f(\beta^\top y_{t-1})\right]$. Then under H_0, it follows

$$E\left[u_{it} E[u_{it} | \beta^\top y_{t-1}] f(\beta^\top y_{t-1})\right] = 0, \tag{3.12}$$

since $E[u_{it} | \beta^\top y_{t-1}] = 0$. Under H_1, we have $E[u_{it} | \beta^\top y_{t-1}] = F_i(\beta^\top y_{t-1}) - \alpha_i \beta^\top y_{t-1}$. Using the law of iterated expectations, we obtain under H_1

$$E\left[u_{it} E[u_{it} | \beta^\top y_{t-1}] f(\beta^\top y_{t-1})\right] > 0. \tag{3.13}$$

Due to (3.12) and (3.13) it is obvious to use the sample analogue of $E\left[u_{it}E[u_{it}|\beta^\top y_{t-1}]f(\beta^\top y_{t-1})\right]$ as the test statistic. The outer expected value is replaced by its mean, the inner expected value by the Nadaraya-Watson estimator

$$\hat{E}(u_{it}|\beta^\top y_{t-1}) = \frac{1}{(T-1)h}\sum_{j=1 j\neq t}^{T} K\left(\frac{\beta^\top y_{t-1} - \beta^\top y_{j-1}}{h}\right) u_{ij}/\hat{f}(\beta^\top y_{t-1}), \tag{3.14}$$

the density function $f(\cdot)$ by the kernel density estimator (3.6) and the residuals u_{it} by the empirical residuals under the null hypothesis, i.e. $\tilde{u}_{it} = \Delta y_{it} - \hat{\alpha}_i\beta^\top y_{t-1}$. Taking the lagged dependent values into account we substitute for $\tilde{u}_{it}$ the residuals $\hat{u}_{it} = \Delta y_{it} - \hat{\Gamma}_i^{\text{OLS}}\xi_{t-1} - \hat{\alpha}_i\beta^\top y_{t-1}$, where $\hat{\Gamma}_i^{\text{OLS}}$ denotes the estimator of the i-th row of Γ given by (3.8) and $\hat{\alpha}_i$ is the estimator of the i-th row of α under the null hypothesis. Thus, the test statistic is of the form

$$I_i := \frac{1}{T(T-1)h}\sum_{t=1}^{T}\sum_{j=1 j\neq t}^{T} K\left(\frac{\beta^\top y_{t-1} - \beta^\top y_{j-1}}{h}\right)\hat{u}_{it}\hat{u}_{ij}, \quad i = 1, \ldots, p$$

To derive the asymptotic distribution, it is important to note that I_i is a degenerate, second-order U-statistic. Combining the ideas of Fan and Li (1999b) and Li and Wang (1998), it can be shown that I_i is asymptotically normally distributed by applying a central limit theorem for U-statistics of β-mixing processes. Furthermore,

$$\hat{\sigma}_i^2 := \frac{2}{T(T-1)h}\sum_{t=1}^{T}\sum_{j=1 j\neq t}^{T} K^2\left(\frac{\beta^\top y_{t-1} - \beta^\top y_{j-1}}{h}\right)\hat{u}_{it}^2\hat{u}_{ij}^2, \quad i = 1, \ldots, p$$

is a consistent estimator for σ_i^2, the asymptotic variance of $Th^{1/2}I_i$. It is well known that the convergence speed to the normal distribution is quite low. Therefore, bootstrap methods are suggested to approximate the finite sample distribution, see e.g. Li and Wang (1998). Due to the enormous sample size it seems reasonable to rely on the asymptotic approximation given through the asymptotic distribution.

3.5 Results

We present the results in two steps. The starting point is the linear benchmark case. We then proceed to the partially linear model and also present the results for the test of linearity described in the previous section.

3.5.1 Linear error correction model

The following table shows the estimation results of the linear error correction model

$$\begin{aligned} r_t^F &= \mu^F + \sum_{i=1}^{20} \gamma_{1i}^F r_{t-i}^F + \sum_{i=1}^{20} \gamma_{1i}^X r_{t-i}^X + \alpha^F (p_{t-1}^X - p_{t-1}^F) + \epsilon_t^F \\ r_t^X &= \mu^X + \sum_{i=1}^{20} \gamma_{2i}^X r_{t-i}^X + \sum_{i=1}^{20} \gamma_{2i}^F r_{t-i}^F + \alpha^X (p_{t-1}^X - p_{t-1}^F) + \epsilon_t^X, \end{aligned}$$

where p denotes the log prices and r denotes a log return. The index X identifies variables and coefficients relating to the spot market (X, Xetra), the index F identifies variables (adjusted by a discount factor according to the cost-of-carry relation) and coefficients relating to the futures market. The cointegrating vector is pre-specified to $(1, -1)^\top$. The model is estimated by OLS with 20 lags, but to save space we present only the coefficients for lags 1-4. Standard errors are based on the heteroskedasticity-robust covariance estimator. The model is estimated based on quote midpoints and 100188 observations. The results are given in Table 3.3.

Considering the short-run dynamics first, we find that the DAX returns depend negatively on their own lagged values but depend positively on lagged futures returns. Returns in the futures markets exhibit a similar pattern. There is one exception, however, as the coefficient on the first lag of the futures returns is positive and significant. The results of F-tests (not shown in the table) indicate that there is bivariate Granger causality.

Table 3.3: Estimation results of the linear ECM

	XDAX		FDAX	
	Estimates	t-statistic	Estimates	t-statistic
Constant	3.385E-6	4.95	-4.427E-6	-3.80
EC	-0.0087	-14.85	0.0047	5.42
XDAX(-1)	-0.0876	-16.36	0.0542	7.36
XDAX(-2)	-0.0773	-16.22	0.0534	7.83
XDAX(-3)	-0.0632	-14.80	0.0573	7.69
XDAX(-4)	-0.0522	-12.14	0.0489	6.76
FDAX(-1)	0.2107	68.32	0.0358	7.97
FDAX(-2)	0.1572	58.18	-0.0166	-3.81
FDAX(-3)	0.1215	46.31	-0.0173	-3.97
FDAX(-4)	0.0989	37.38	-0.0079	-1.78
R^2	0.2244		0.0070	

The coefficients on the error correction term have the expected signs (negative for the spot market and positive for the futures market) and are both highly significant. The estimates can be used to construct the common factor weights

$$\theta^X = \frac{\alpha^F}{\alpha^F - \alpha^X}; \quad \theta^F = \left(1 - \theta^X\right) = \frac{-\alpha^X}{\alpha^F - \alpha^X}$$

The common factor weights measure the contributions of the two markets to the process of price discovery. The measure builds on Gonzalo and Granger (1995) and is discussed in more detail in Booth et a. (2002), deB Harris et al. (2002) and Theissen (2002). In our linear error correction model the common factor weights are 0.3507 for the spot market and 0.6493 for the futures market. The futures market thus dominates in the process of price discovery. This result is consistent with previous findings.

3.5.2 Partially linear error correction model

Applying the test for linearity developed in section 3.4, we obtain $I^F = 3.265$ and $I^X = 2.937$. We thus clearly reject the linear benchmark model in favor of our non-parametric specification. For the test we choose the bandwidth parameter to be $h = 2\hat{\sigma}T^{-0.2}$. The following table shows the estimation results of the partially linear error correction model

$$
\begin{aligned}
r_t^F &= \sum_{i=1}^{20} \gamma_{1i}^F r_{t-i}^F + \sum_{i=1}^{20} \gamma_{1i}^X r_{t-i}^X + F(p_{t-1}^X - p_{t-1}^F) + \epsilon_t^F \\
r_t^X &= \sum_{i=1}^{20} \gamma_{2i}^X r_{t-i}^X + \sum_{i=1}^{20} \gamma_{2i}^F r_{t-i}^F + F(p_{t-1}^X - p_{t-1}^F) + \epsilon_t^X,
\end{aligned}
$$

where the notation is as in the linear model. We estimate the model by the procedure described in section 3.3. Again, we use 20 lags, but only the coefficients for lags 1-4 are shown. Again, standard errors are based on the heteroskedasticity-robust covariance estimator. The cointegrating vector is pre-specified to $(1, -1)^\top$.

Table 3.4: Estimation results of the partially linear ECM ($h = 2\hat{\sigma}T^{-0.2}$)

	XDAX		FDAX	
	Estimates	t-statistic	Estimates	t-statistic
XDAX(-1)	-0.0873	-15.25	0.0389	4.79
XDAX(-2)	-0.0693	-14.90	0.0475	6.15
XDAX(-3)	-0.0564	-13.57	0.0491	5.78
XDAX(-4)	-0.0435	-10.76	0.0449	5.54
FDAX(-1)	0.1571	70.98	0.0558	11.39
FDAX(-2)	0.1351	58.79	0.0020	0.39
FDAX(-3)	0.1063	47.14	-0.0053	-1.05
FDAX(-4)	0.0882	39.27	-0.0028	-0.54

The results for the short-run dynamics are similar to those in the linear model. The spot market returns depend positively on their own lagged values and negatively on the lagged futures returns. Futures returns, on the other hand, depend positively on the lagged spot market returns. They also depend positively on their first lag. Coefficients for higher lags are insignificant.

Figure 3.1 presents the results for the adjustment process. The figure plots the value of the adjustment function F against the pricing error $\beta^\top y_{t-1}$.

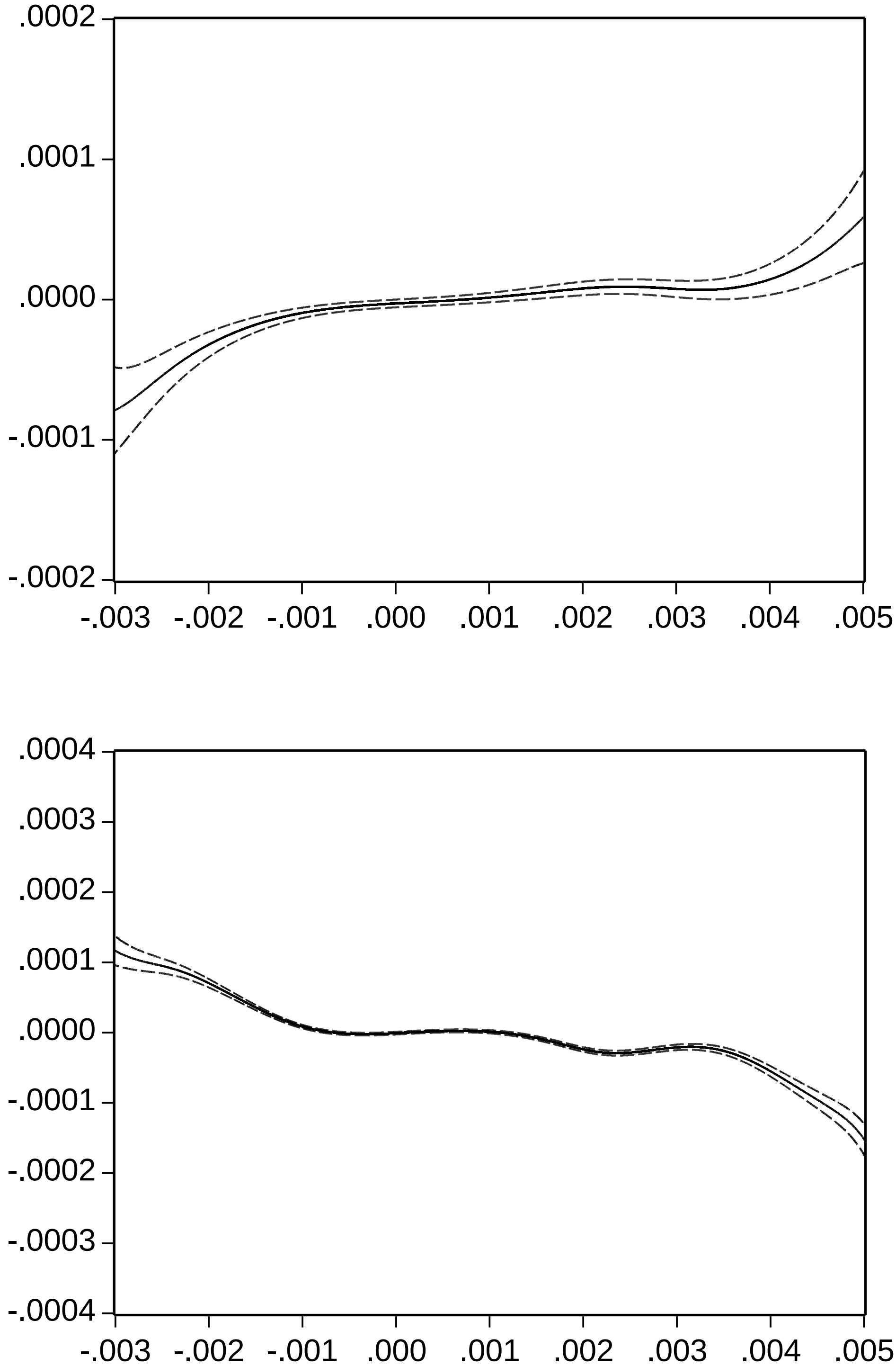

Figure 3.1: Estimated adjustment process (solid line) and pointwise 95% confidence interval (dashed line) for FDAX (upper panel) and XDAX (lower panel) as a function of the error correction term $p^X - p^F$. A Gaussian kernel and the bandwidths $h^F = 0.000492$ and $h^X = 0.000361$ have been used.

It also depicts the 95% confidence intervals. The upper panel shows the results for the futures market, the lower panel those for the spot market.The adjustment process is estimated very precisely, as evidenced by the narrow confidence intervals. In the outer regions (i.e., when pricing errors are large) estimation is less precise. This is a natural consequence of the low number of observations in these regions.

The speed of adjustment is almost monotonically related to the magnitude of the pricing error. This shape of the adjustment function is clearly at odds with a threshold error correction model. Adjustment is slow for small pricing errors, as is evidenced by the small slope of the adjustment function. When the pricing error becomes larger, the speed of adjustment increases sharply. This is consistent with arbitrage activities.

There is an asymmetry with respect to the level of the pricing error that triggers arbitrage. When the pricing error is negative (i.e., when the adjusted futures price is larger than the spot price) the trigger level is about -0.001. When the pricing error is positive, on the other hand, the trigger level is approximately 0.003. This pattern is explained by slight, but systematic deviations of prices from the cost-of-carry relation. On average, the difference between the discounted futures price and the DAX index is -2.8 index points. This pattern has been documented in previous research (e.g. Bühler and Kempf, 1995), and the most likely explanation is differential tax treatment of dividends in the spot and the futures market (see McDonald, 2001 for a detailed discussion).

In order to compare the predictive ability of the partially linear VECM with that of the linear VECM, the root mean squared error (RMSE) and the mean absolute error (MAE) are calculated for both models.[6] The RMSE and the MAE are defined

[6] We restrict the analysis of the forecasting errors to the XDAX equation. This equation lends itself to forecasting because of the large and significant coefficients on the lagged futures returns documented in table 3.4.

for one-step ahead forecast errors by

$$
\begin{aligned}
RMSE &= \sqrt{\sum_{t=k}^{T} \left(\hat{E}_{t-1} p_t^X - p_t^X \right)^2}, \\
MAE &= \sum_{t=k}^{T} \left| \hat{E}_{t-1} p_t^X - p_t^X \right|.
\end{aligned}
$$

We set $k = 80000$ to ensure that the parameter estimates are based on a sufficiently large numbers of observations. The results are given in Table 3.5.

Table 3.5: Prediction ability of the linear VECM and the partially linear VECM

	Linear VECM (A)	Partially Linear VECM (B, B/A)
RMSE	0.025	0.023 (0.919)
MAE	2.276	2.067 (0.908)

Table 3.5 shows that the root mean squared error (RMSE) of the partially linear VECM is about 10% lower than the linear VECM. A similar result is obtained for the mean absolute error (MAE). Hence, the partially linear VECM clearly improves the forecasting ability.

3.6 Conclusion

The present chapter extends the literature on the joint dynamics of prices in spot and futures markets by modelling the price-adjustment process non-parametrically using the methodology developed in Gaul (2005).

We apply our partially linear error correction model to data for the German blue chip index DAX and the DAX futures contract traded on the EUREX. We find that the adjustment process is indeed nonlinear. The linear benchmark case is rejected at all reasonable levels of significance. Consistent with economic intuition, the speed of adjustment is almost monotonically increasing in the magnitude of the pricing error (the deviation between discounted futures price and spot price). This pattern is inconsistent with a simple threshold error correction model. It is consistent with a smooth transition model, and in fact the shape of the adjustment process in our non-parametric model may guide the choice of the transition function in future empirical research.

Bibliography

[1] Anderson, H. M. and F. Vahid (2001), Market Architecture and Nonlinear Dynamics of Australian Stock and Futures Indices, *Australian Economic Papers* **40**, 541-566.

[2] Andrews, D. W. K. (1993), Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica* **61**, 821-856.

[3] Andrews, D. W. K. (1994), Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity, *Econometrica* **62**, 43-72.

[4] Andrews, D. W. K. and W. Ploberger (1994), Optimal Tests when a Nuisance Parameter Is Present only under the Alternative, *Econometrica* **62**, 1383-1414.

[5] Bai, J. (1997), Estimating Multiple Breaks One at a Time, *Econometric Theory* **13**, 315-352.

[6] Balke, N. S. and T. B. Fomby (1997), Threshold Cointegration, *International Economic Review* **38**, 627-645.

[7] Beare, B. K. (2006), Unit Root Testing with Unstable Volatility, mimeo, Yale University.

[8] Bec, F. and A. Rahbek (2004), Vector Equilibrium Correction Models with Nonlinear Discontinuous Adjustments, *Econometrics Journal* **7**, 628-651.

[9] Berkmann, H., Brailsford, T. and A. Frino (2005), A Note on Execution Costs for Stock Index Futures: Information Versus Liquidity Effects, *Journal of Banking and Finance* **29**, 565-577.

[10] Booth, G., Lin, J., Martikainen, T. and Y. Tse (2002) Trading and Pricing in Upstairs and Downstairs Markets, *Review of Financial Studies* **15**, 1111-1135.

[11] Boswijk, H. P. (2005), Adaptive Testing for a Unit Root with Nonstationary Volatility, UvA-Econometrics Discussion Paper 2005/07.

[12] Buehler, W. and A. Kempf (1995), DAX Index Futures: Mispricing and Arbitrage in German Markets, *Journal of Futures Markets* **15**, 833-859.

[13] Campbell, J. and R. Shiller (1987), Cointegration and Tests of Present Value Models, *Journal of Political Economy* **95**, 1062-1088.

[14] Caner, M. and B. E. Hansen (2001), Threshold Autoregression with a Unit Root, *Econometrica* **69**, 1555-1596.

[15] Cavaliere, G. (2004), Unit Root Tests under Time-varying Variances, *Econometric Reviews* **23**, 259-292.

[16] Cavaliere, G. and A. M. R. Taylor (2008), Bootstrap Unit Root Tests for Time Series with Nonstationary Volatility, *Econometric Theory* **24**, 43-71.

[17] Chan, K. S. (1993), Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *Annals of Statistics* **21**, 520-533.

[18] Davies, R. B. (1987), Hypothesis Testing when a Nuisance Parameter Is Present only under the Alternative, *Biometrika* **74**, 33-43.

[19] deB Harris, F., McInish, T. and R. Wood (2002), Common factor components versus information shares: A reply, *Journal of Financial Markets* **5**, 341-348.

[20] Dickey, D. A. and W. A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* **74**, 427-431.

[21] Dwyer Jr., G. P., Locke, P. and W. Yu (1996), Index Arbitrage and Nonlinear Dynamics between the S&P 500 Futures and Cash, *Review of Financial Studies* **9**, 301-332.

[22] Elliott, G., Rothenberg, T. J. and J. H. Stock (1996), Efficient Tests for an Autoregressive Unit Root, *Econometrica* **64**, 813-836.

[23] Engle, R. F. and C. W. J. Granger (1987), Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica* **55**, 251-276.

[24] Fan, J. and Q. Yao (2003), *Nonlinear Time Series - Nonparametric and Parametric Methods*, Springer Series in Statistics, New York and Heidelberg.

[25] Fan, Y. and Q. Li (1999a), Root-N-Consistent Estimation of Partially Linear Time Series Models, *Journal of Nonparametric Statistics* **11**, 251-269.

[26] Fan, Y. and Q. Li (1999b), Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing, *Journal of Nonparametric Statistics* **10**, 245-271.

[27] Gaul, J. (2005), A Partially Linear Vector Error Correction Model, Unpublished Manuscript, University of Bonn.

[28] Gaul, J. (2007), Estimation and Testing in a Three-Regime Vector Error Correction Model, Unpublished Manuscript, University of Bonn.

[29] Gaul, J. (2008), Unit Root Tests with Time-Varying Variances - A Simulation Study, Unpublished Manuscript, University of Bonn.

[30] Gaul, J. and E. Theissen (2007), A Partially Linear Approach to Modelling the Dynamics of Spot and Futures Prices, Unpublished Manuscript, University of Bonn.

[31] Gonzalo, J. and J. Pitarakis (2006), Threshold Effects in Multivariate Error Correction Models, in T.C. Mills and K. Patterson (eds), *Palgrave Handbook of Econometrics: Econometric Theory* **Volume I, Chapter 15**, 578-609.

[32] Gonzalo, J. and C. W. J. Granger (1995), Estimation of Common Long-Memory Components in Cointegrated Systems, *Journal of Business and Economic Statistics* **13**, 27-35.

[33] Härdle, W. (1991), *Smoothing Techniques - With Implementation in S*, Springer: New York.

[34] Härdle, W., Müller, M., Sperlich, S. and A. Werwatz (2004), *Nonparametric and Semiparametric Models*, Springer: New York.

[35] Härdle, W. and D. W. Scott (1992), Smoothing in by Weighted Averaging Using Rounded Points, *Computational Statistics* **7**, 97-128.

[36] Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, New Jersey.

[37] Hamori, S. and A. Tokihisa (1997), Testing for a Unit Root in the Presence of a Variance Shift, *Economics Letters* **57**, 245-253.

[38] Hansen, B. E. (1996), Inference when a Nuisance Parameter Is not Identified under the Null Hypothesis, *Econometrica* **64**, 413-430.

[39] Hansen, B. E. (2000), Sample Splitting and Threshold Estimation, *Econometrica* **68**, 575-603

[40] Hansen, B. E. and B. Seo (2002), Testing for Two-Regime Threshold Cointegration in Vector Error Correction Models, *Journal of Econometrics* **110**, 293-318.

[41] Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, *Journal of Economics Dynamic and Control* **12**, 231-254.

[42] Johnasen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica* **59**, 1551-1580.

[43] Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

[44] Kawaller, I. G., Koch, P. D. and T. W. Koch (1987), The Temporal Price Relationship between S&P 500 Futures and the S&P 500 Index, *Journal of Finance* **42**, 1309-1329.

[45] Kim,T. H., Leybourne, S. and P. Newbold (2002), Unit Root Tests with a Break in Innovation Variance, *Journal of Econometrics* **59**, 365-387.

[46] Li, Q. and S. Wang (1998), A Simple Consistent Bootstrap Test for a Parametric Regression Function, *Journal of Econometrics* **87**, 145-165.

[47] Lo, M. C. and E. Zivot (2001), Threshold Cointegration and Nonlinear Adjustment to the Law of One Price, *Macroeconomic Dynamics* **5**, 533-576.

[48] Lütkepohl, H. (2006), *New Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin.

[49] Mankiw, G. and J. Miron (1986), The Changing Behavior of the Term Structure of Interest Rates, *Quarterly Journal of Economics* **110**, 211-228.

[50] Martens, M., Kofman, P. and T. C. F. Vorst (1998), A Threshold Error-Correction Model for Intraday Futures and Index Returns, *Journal of Applied Econometrics* **13**, 245-263.

[51] Masry, E. and D. Tjostheim (1995), Nonparametric Estimation and Identification of Nonlinear ARCH Time Series: Strong Convergence and Asymptotic Normality, *Econometric Theory* **11**, 258-289.

[52] McDonald, R. L. (2001), Cross-Border Investing with Tax Arbitrage: The Case of German Dividend Tax Credits, *Review of Financial Studies* **14**, 617-657.

[53] Mokkadem, A. (1988), Mixing Properties of ARMA Processes, *Stochastic Processes Appl.* **29**, 309-315.

[54] Ng, S. and P. Perron (2001), Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, *Econometrica* **69**, 1519-1554.

[55] Perron, P. (1989), The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, *Econometrica* **57**, 1361-1401.

[56] Perron, P. (1990), Testing for a Unit Root in a Time Series with a Changing Mean, *Journal of Business and Economic Statistics* **8**, 153-162.

[57] Phillips, P. C. B. and J. C. Chao (1999), Model Selection in Partially Nonstationary Vector Autoregressive Processes with Reduced Rank Structure, *Journal of Econometrics* **91**, 227-271.

[58] Phillips, P. C. B. and S. Durlauf (1986), Multiple Time Series Regression with Integrated Processes, *Review of Economic Studies* **53**, 473-495.

[59] Phillips, P. C. B. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, *Biometrika* **75**, 335-346.

[60] Robinson, P. (1988), Root-N Consistent Semiparametric Regression, *Econometrica* **56**, 931-954.

[61] Roll, R., Schwartz, E. and A. Subrahmanyam (2007), Liquidity and the Law of One Price: The Case of the Futures-Cash Basis, *Journal of Finance* **62**, 2201-2234.

[62] Seo, M. (2004), Bootstrap Testing for the Presence of Threshold Cointegration in a Threshold Vector Error Correction Model, Unpublished Manuscript, Department of Economics, University of Wisconsin-Madison.

[63] Sensier, M. and D. van Dijk (2004), Testing for Volatility Changes in U.S. Macroeconomic Time Series, *Review of Economics and Statistics* **86**, 833-839.

[64] Shyy, G., Vijayraghavan, V. and B. Scott-Quinn (1996), A Further Investigation of the Lead-Lag Relationship between the Cash Market and Stock Index Futures Market with the Use of Bid/Ask Quotes: The Case of France. *Journal of Futures Markets* **16**, 405-420.

[65] Stock, J. H. (1999), A Class of Tests for Integration and Cointgeration, in R. F. Engle and H. White (eds), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, 137-167, Oxford University Press.

[66] Stoll, H. and R. Whaley (1990), The Dynamics of Stock Index and Stock Index Futures Returns, *Journal of Financial and Quantitative Analysis* **25**, 441-468.

[67] Taylor, N., van Dyck, D., Franses, P. H. and A. Lucas (2000), SETS, Arbitrage Activity and Stock Price Dynamics, *Journal of Banking and Finance* **24**, 1289-1306.

[68] Theissen, E. (2002), Price Discovery in Floor and Screen Trading Systems, *Journal of Empirical Finance* **9**, 455-474.

[69] Theissen, E. (2005), Price Discovery in Spot and Futures Markets: A Reconsideration. Unpublished Manuscript. University of Bonn.

[70] Tse, Y. (2001) Index Arbitrage with Heterogenous Investors: A Smooth Transition Error Correction Analysis, *Journal of Banking and Finance* **25**, 1829-1855.

[71] van Dijk, D. and P. H. Franses (2000), Nonlinear Error-Correction Models for Interest Rates in the Netherlands, in *Nonlinear Econometric Modelling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory*, Barnett, W. A., Hendry, D., Hylleberg, S., Teräsvirta, T., Tjostheim, D. and A. Wurtz (eds), Cambridge University Press: Cambridge, 203-227.

[72] Yadav, P. K., Pope, P. F. and K. Paudyal (1994), Threshold Autoregressive Modelling in Finance: The Price Differences of Equivalent Assets, *Mathematical Finance* **4**, 205-221.

[73] Zheng, J. X. (1996), A Consistent Test of Functional Form via Nonparametric Estimation Techniques, *Journal of Econometrics* **75**, 263-289.