Essays on Persistence in Economic Time Series

Von der Wirtschaftswissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

– Doctor rerum politicarum –

genehmigte Dissertation

von

Diplom-Volkswirt Robinson Kruse

geboren am 25. Oktober 1981 in Wien

2009

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Koreferent: Prof. Dr. Jörg Breitung, Universität Bonn

Tag der Promotion: 14.11.2008

Nichts ist getan, wenn noch etwas zu tun übrig ist.

– Carl Friedrich Gauß

Acknowledgements

I am grateful to my advisor and co-author Philipp Sibbertsen. He supported me strongly, gave me a lot of liberties and confidence. Moreover, I was able to benefit from his research expertise, especially in the field of long memory time series processes. I have to thank all my colleagues at the Institute of Statistics, especially Gudrun Westphal, for providing me an enjoyable working atmosphere.

Moreover, I am grateful to Jörg Breitung for all of his valuable advices, efforts and support. I am indebted to my other co-authors Michael Frömmel and Lukas Menkhoff and I would like to thank Jürgen Wohlers and his team from the computer lab (ITS-Pool) for generous technical assistance.

My friends at the universities of Berlin, Hannover and Leipzig, Marion Berlauer, Chris Hecker, Falko Tabbert and Christina Ziegler, also deserve a great deal of gratitude.

As in all my endeavors, Alexandra Becher was very supportive and I am deeply grateful to her.

Hannover, July 2008
Robinson Kruse

Kurzfassung

Die Analyse von Persistenzeigenschaften ökonomischer Zeitreihen hat wegen ihrer Bedeutung für viele wirtschaftliche Aspekte eine lange Tradition in der Ökonometrie. Diese Arbeit beinhaltet fünf Essays die sich mit einer Vielzahl von Themen im Rahmen der modernen Modellierung von Persistenz beschäftigen. Darunter befinden sich Einheitswurzeln, langfristige Abhängigkeit, Strukturbrüche und Nichtlinearitäten.

Kapitel 2 wurde zusammen mit Philipp Sibbertsen verfasst und untersucht das Inferenzproblem eines Strukturbruchs im fraktionalen Integrationsgrad einer Zeithreihe. Es wird ein modifizierter Test vorgeschlagen und die asymptotischen Eigenschaften sowie das Verhalten in kleinen Stichproben analysiert. Im dritten Kapitel wird dieser Test angewendet um die Hypothese einer rationalen Blase im Standard and Poors 500-Aktienindex empirisch zu überprüfen. Die Resultate lassen neue Schlussfolgerungen über die Existenz langfristiger Abhängigkeiten und die Präsenz von Strukturbrüchen zu. Ein neuer Test für die Einheitswurzelhypothese gegen die Alternative eines populären nichtlinearen Zeitreihenmodells wird in Kapitel 4 vorgeschlagen. Der neue Test verallgemeinert einen bislang häufig verwendeten Test durch den Einsatz einer neuen Inferenztechnik und ist diesem durch eine höhere Güte überlegen.

Das fünfte Kapitel ist eine Zusammenarbeit mit Michael Frömmel, Lukas Menkhoff und Philipp Sibbertsen und untersucht das Problem der empirischen Falsifizierbarkeit der Kaufkraftparität durch den Einsatz nichtlinearer Einheitswurzeltests unter Bedingungen, die in der Praxis vorherrschen. Die empirischen Ergebnisse deuten darauf hin, dass Markov-Switching Prozesse die Hypothese der Kaufkraftparität stützen. Im letzten Kapitel, dass mit Philipp Sibbertsen verfasst wurde, wird ein dominantes Verfahren zur Modellselektion für potenziell nichtlineare und nichtstationäre Modelle vorgeschlagen.

Schlagwörter: Einheitswurzeln, langes Gedächtnis, Strukturbrüche, Nichtlinearitäten

Short summary

The analysis of persistence properties of economic time series has a long tradition in econometrics due to its paramount importance for many economic issues. This collection of five essays deals with a variety of issues in modern persistence modeling. Among these are unit roots, long-range dependence, structural breaks and non-linearity.

Chapter 2, written together with Philipp Sibbertsen, considers the inference problem of a structural break in the fractional degree of integration. A modified test is proposed and its asymptotic and small sample behaviour is studied. In chapter 3, the test is applied to the problem of testing for a bubble in the Standard and Poors 500 stock market index. New results on long-range dependence and structural change are obtained. A new test for the unit root hypothesis against a popular non-linear time series model is proposed in chapter 4. The new test generalizes an extant test by making use of a new non-standard inference technique. Numerical results suggest that the new test is generally superior in terms of power.

Chapter 5, co-authored with Michael Frömmel, Lukas Menkhoff and Philipp Sibbertsen, considers the problem of falsifying Purchasing Power Parity empirically by using non-linear unit root tests under conditions that are relevant in practice. The empirical results suggest that Markov Switching processes which include the modeling of destabilizing forces in foreign exchange rates support the Purchasing Power Parity hypothesis. Lastly, chapter 6, written together with Philipp Sibbertsen, deals with the decision problem regarding four different types of time series processes. A dominant model selection strategy for potentially non-linear and non-stationary models is suggested.

Keywords: Unit roots, long memory, structural breaks, non-linearity

Contents

Chapter 1

Introduction

The analysis of persistence properties of economic time series has a long tradition in econometrics. A great deal of literature has focused on linear processes and their persistence structures. Stationary processes and deterministic trends have especially been considered until the outset of unit roots which have become one of the most important research issues in modern time series analysis. The growing interest in unit roots can be explained by the fact that trending is one of the most dominant characteristics of economic time series. Additionally, deterministic trend processes are very limited. On the contrary, unit root processes are stochastically trending and hence imply the permanency of shocks to economic variables. Such behavior is called persistent.

Without a doubt, the persistence properties of economic time series are of paramount importance for many economic issues. Policy makers for instance have to know how shocks affect certain variables in the short and the long run. Economic forecasting builds upon time series models that reflect the persistence properties of the underlying variables. Moreover, the persistence of shocks is of ultimate importance for testing economic theories like Purchasing Power Parity (PPP). PPP holds if and only if the real exchange rate follows a stationary process which rules out any unit roots. In other words, shocks to the real exchange rate have to be transitory. There are numerous examples of economic theories that can be falsified by testing for unit roots.

This collection of five essays deals with different perspectives on persistence. The common theme of all the essays is statistical inference for univariate time series processes. Extensions can also be made into multivariate processes and dynamic panel data models. The two main differences in this work are fractional integration and non-linearity. The former concept resolves the classical paradigm of an integer degree of integration that is typically zero or one for economic variables. Allowing for a fractional instead of an integer degree of integration implies increased modeling flexibility and more importantly, long-range dependence of shocks. Hence, the class of fractionally integrated time series models offers a different view on the persistence. A synonym for fractional integration is long memory, as shocks have a long lasting impact. Furthermore, long memory time series models do not only have theoretical appeal. There are a lot of empirical studies, including those outside the filed of economics, that successfully apply them to a variety of problems and types of variables.

Chapter 2, written together with Philipp Sibbertsen, considers the inference problem of a structural break in the fractional degree of integration. Leybourne et al. (2007) proposed a CUSUM of squares test for the unit root hypothesis against the alternative that the integer degree of integration changes from zero, which implies stationarity, to one, which implies non-stationarity, at some breakpoint in time. This test is generalized with respect to fractional integration. Several new theoretical results are given and the problem of conservatism that is inherent in the original test by Leybourne et al. (2007) is resolved. The small sample performance of the modified test is analyzed by means of a Monte Carlo study and it appears to work well. An application to the US inflation rate shows the empirical relevance of a break in the persistence of long memory models.

Chapter 3 is dedicated to the problem of testing for a bubble in the Standard and Poors 500 (S&P 500) stock market index. This application is motivated by two articles analyzing the persistence structure of the logarithm of dividend yields. Sollis (2006) finds a change in persistence by using methods for integer integration. These results indicate a rational bubble in the stock market. This is in contrast to the results reported in Koustas and Serletis (2005) which suggest that the series is fractionally integrated but there is no evidence to indicate a rational bubble. However, Koustas and Serletis (2005) do not take potential structural breaks into account. Therefore, it is suitable to apply the modified CUSUM of squares test introduced in chapter 2 to this data set. The results highlight two empirical findings: on the one hand they confirm the previous result of fractional integration and on the other hand they support the hypothesis of a rational bubble.

Another viewpoint on persistence is the one implied by non-linear regime switching time series processes. The main idea is that the data generating process exhibits more than one regime or state of nature. Linear models implicitly assume that only one state exists and hence, there cannot be any switch of certain characteristics in the data generating processes over time. Structural breaks are understood as deterministic regime shifts. When non-linear models are applied in economics, it is quite common to assume that there are two or three regimes where every regime has different parameters and conditions describing the data generating mechanism. This has an immediate consequence on persistence since it can now be interpreted in a local and a global sense. Local persistence means the persistence of the time series process in a certain regime, while global persistence describes the overall persistence. Note that local and global persistence are the same for linear models. To put it differently, non-linear time series models allow for time-varying instead of constant persistence.

One of the major issues with regard to the family of regime switching models, which comprises Markov switching and smooth transition autoregressive (STAR) models for instance, is the determination and speed of a switch between regimes. STAR models assume a smooth transition between regimes that is usually determined by observable past values of the process itself or by the time period. In the latter case, the switch becomes deterministic. In Markov switching models regime changes are determined by an unobservable stochastic Markov process. However, the popularity of non-linear models in applied economics and econometrics is due to their convenient interpretation in the context of economic models and their increased flexibility in comparison to linear models. A major drawback of non-linear time series models is complicated inference. This is because certain parameters are usually not identified under the null hypothesis. Therefore, non-standard methods have to be applied. Alternatively, and this is done more often, the non-linear model is linearized if possible, which is of particular interest for smooth transition models. Standard techniques can be used in the linearized model, which is a clear advantage but such an approximation may waste important information. Since the Markov switching model cannot be linearized, inference is much more involved.

A new test for the unit root hypothesis against a stationary exponential STAR model is proposed in chapter 4. The new test generalizes the extant test by Kapetanios et al. (2003) that makes use of linearization and builds upon a new non-standard inference technique suggested by Abadir and Distaso (2007). First, an empirically unrealistic assumption about the location parameter of the smooth transition function is relaxed. Second, the resulting auxiliary regression implies a non-standard testing problem in the sense that one parameter is one-sided under the alternative while all others are two-sided. This gives rise to the modified test statistics introduced by Abadir and Distaso (2007) that are explicitly designed for such problems. The limiting distribution of the modified Wald test is derived and consistency is proven. Using Monte Carlo simulations the extant and the new test are compared under a variety of conditions. Overall, the new test appears to be superior in many situations. An empirical application to the EU real effective exchange rate underlines its usefulness. PPP is supported by the new test while it is rejected by the existing one and two prominent linear unit root tests.

Chapter 5, co-authored with Michael Frömmel, Lukas Menkhoff and Philipp Sibbertsen, considers the problem of falsifying PPP empirically by using non-linear unit root tests under conditions that are relevant in practice. In particular, this means that sample sizes and parameter settings are chosen carefully in the conducted large-scale Monte

Carlo study. Beside the two non-linear unit root tests against exponential STAR and the famous Dickey-Fuller test against linear autoregressive models, a newly proposed unit root test against Markov switching autoregression is studied. This new test has a stationary two regime Markov switching model as the alternative and is constructed upon the methodology for unidentified parameters under the Null that was proposed in Hansen (1996) and has subsequently been refined by Garcia (1998). It appears to be relatively powerful and robust against non-linear ESTAR-type dynamics. The simulation results show that these two important properties are not shared by other studied tests. This suggests that Markov Switching processes which include the modeling of destabilizing forces in foreign exchange rates may be appropriate.

Lastly, chapter 6, written together with Philipp Sibbertsen, deals with the decision problem regarding four different types of time series processes. The first two types are stationary and non-stationary processes, respectively. As mentioned above, it is of great interest for several reasons to distinguish these two types. It is widely acknowledged that non-linearities play an important role in economics, especially in macroeconomics and finance. Therefore, it is relevant to discriminate between linear and non-linear processes, too. Up to now, both strands of the literature are usually isolated. In a recent article, Harvey and Leybourne (2007) use a stationarity test that claimed to be robust against non-linearity and a linearity test that is robust against non-stationarity. This was done in order to classify European real exchange rates into the four possible categories. Firstly, the quality of the proposed decision rule, which uses two tests simultaneously, is investigated. Secondly, some alternative strategies and modifications are suggested and compared to the original one. A broad simulation study shows that sequential procedures outperform simultaneous ones. It turns out that no dominating rule exists since every decision rule has its own advantages and disadvantages. Fortunately, it is possible to construct a dominant strategy via a simple pre-testing strategy and the unification of two well performing procedures.

Chapter 2

Testing for a Break in Persistence under Long-Range Dependencies

Co-authored with Philipp Sibbertsen. Published as Hannover Leibniz University Discussion Paper No. 381, Nov. 2007 (ISSN 0949-9962) and as a revision in the Journal of Time Series Analysis (2009), Issue 30, pp. 263-285.

2.1 Introduction

For a practitioner it is of big importance in terms of model building and forecasting to know whether a given time series has a certain kind of persistence, either stationary $I(d)$ with $0 \leq d < 1/2$ or non-stationary $I(d)$ with $1/2 < d < 3/2$ or whether the persistence breaks from stationary to non-stationary persistence or vice versa. Recently, a number of tests for a break in the persistence have been proposed in the classical $I(0)/I(1)$ framework. Kim (2000), Kim et al. (2002) and Busetti and Taylor (2004) propose tests for the null hypothesis that the data generating process is $I(0)$ throughout against the alternative of a break to $I(1)$. Contrary to these tests Banerjee et al. (1992) and Leybourne et al. (2003) propose tests for the opposite null of $I(1)$ throughout against the alternative of a break to $I(0)$. All these tests have problems when the data generating process does not exhibit a break in persistence, i.e. when the null is false as well. Therefore, Leybourne et al. (2007) proposed a CUSUM of squares-based test to overcome this problem. The Leybourne et al. (2007) test is basically the ratio of two

CUSUM of squares statistics, based on the forward and reverse evaluation of the time series. Although the test is constructed for the null hypothesis that the data generating process is $I(1)$ throughout against a break in persistence to $I(0)$, Leybourne et al. show that it has also power against the alternative of a break from $I(0)$ to $I(1)$ and that it behaves well if the process is $I(0)$ throughout.

However, all of these tests stay in the classical $I(0)/I(1)$ framework. One exception is Beran and Terrin (1996) who consider a test for constancy of the long-memory parameter against a change of it. Their test is based on a functional central limit theorem for quadratic forms. By now it is broadly accepted that many economic variables exhibit long-range dependencies which cannot be covered by the classical framework. Also in the more flexible $I(d)$ framework, $0 \leq d \leq 3/2$ it is crucial to know whether the memory parameter is in the stationary region or in the non-stationary region throughout or whether there is a change in the persistence. It turns out that the Leybourne et al. (2007) test has serious size distortions, that means the test is conservative, if the data generating process has long memory and therefore the test has a lack of power in this model. This indicates that new critical values depending on the memory parameter are necessary in the $I(d)$ framework. In this chapter we investigate the asymptotic behavior of the Leybourne et al. (2007) test under long-range dependencies. We derive the limiting distribution under the null that the data generating process exhibits non-stationary long memory. We furthermore show that the breakpoint estimator proposed by Leybourne et al. (2007) is also consistent under long memory though with a slower rate of convergence depending on d. In a Monte Carlo study we show that the test has satisfying size and power properties when the adjusted critical values are used. Finally the test is applied to monthly US inflation data.

The chapter is organized as follows. After introducing the model and the test in section 2.2, section 2.3 derives the asymptotic properties of the test. Section 2.4 contains an intensive Monte Carlo study showing the finite sample properties of the test as well as the power properties and gives response curves to easily compute critical values. Section 2.5 contains an empirical application to a monthly US inflation time series and section

2.6 concludes. All proofs are given in the Appendix A of this chapter, while Appendix B contains additional numerical results.

2.2 Model and Test

We assume that the data generating process follows an ARFIMA(p, d, q) process as proposed by Granger and Joyeux (1980):

$$\Phi(B)(1-B)^d X_t = \Psi(B)\varepsilon_t,$$

where ε_t are i.i.d. random variables with mean zero and variance σ^2. The AR- and MA-polynomials $\Phi(B)$ and $\Psi(B)$ are assumed to have all roots outside the unit circle. The degree of integration of X_t is therefore solely determined by the memory parameter d. The test against a change in the persistence as proposed by Leybourne et al. (2007) uses the statistic

$$R = \frac{\inf_{\tau\in\Lambda} K^f(\tau)}{\inf_{\tau\in\Lambda} K^r(\tau)}, \tag{2.1}$$

where $K^f(\tau)$ and $K^r(\tau)$ are CUSUM of squares-based statistics depending on the forward and reversed residuals of the data generating process as given below. Here τ is the relative breakpoint where we assume that $\tau \in \Lambda$ and that $\Lambda \subset (0, 1)$ is symmetric around 0.5. For now we assume τ to be fixed though unknown. As τ is usually unknown in practice we study the properties of a simple estimator for the breakpoint in the following section. In detail CUSUM of squares-based statistics are defined by

$$K^f(\tau) = [\tau T]^{-2} \sum_{t=1}^{[\tau T]} \hat{v}_{t,\tau}^2$$

and

$$K^r(\tau) = (T - [\tau T])^{-2} \sum_{t=1}^{T-[\tau T]} \tilde{v}_{t,\tau}^2.$$

We denote by $[x]$ the biggest integer smaller than x. Here, $\hat{v}_{t,\tau}$ is the residual from the OLS regression of X_t on a constant $z_t = 1\ \forall t$ based on the observations up to $[\tau T]$. This is

$$\hat{v}_{t,\tau} = X_t - \bar{X}(\tau)$$

with $\bar{X}(\tau) = [\tau T]^{-1} \sum_{t=1}^{[\tau T]} X_t$. Similarly $\tilde{v}_{t,\tau}$ is defined for the reversed series $y_t = X_{T-t+1}$. Thus, it is given by

$$\tilde{v}_{t,\tau} = y_t - \bar{y}(1-\tau)$$

with $\bar{y}(1-\tau) = (T - [\tau T])^{-1} \sum_{t=1}^{T-[\tau T]} y_t$. The case of $z_t = [1,\ t]'$, which corresponds to linear de-trending, is considered later on as well.

Remark: It should be mentioned that the quantities $K^f(\tau)$ and $K^r(\tau)$ are originally defined by including an estimator of the long-run variance of the data generating process. As the behavior of the test statistic R is independent of the long-run variance we distance ourselves from this issue to keep notation and proofs simple.

2.3 Asymptotic Properties

In this section we derive the asymptotic properties of the test statistic (2.1) when the data generating process is $I(d)$. In the following we denote by $\Rightarrow$ weak convergence and by $\xrightarrow{P}$ convergence in probability. We denote by d_0 the long memory parameter under the null hypothesis regardless of its specific value while we distinguish under the alternative hypothesis between values characterizing stationary $(0 \leq d_1 < 1/2)$ and non-stationary processes $(1/2 < d_2 < 3/2)$, respectively.

Theorem 1. *Under the null hypothesis $H_0 : X_t \sim I(d_0)\ \forall t$ with $1/2 < d_0 < 3/2$ the limiting distribution for $T \to \infty$ of R is given by*

$$T^{-2d_0} R \Rightarrow \frac{\inf_{\tau \in \Lambda} L_{d_0}^f(\tau)}{\inf_{\tau \in \Lambda} L_{d_0}^r(\tau)}$$

with

$$
\begin{aligned}
L^f_{d_0}(\tau) &= \int\limits_0^{\tau} W^*_{d_0}(r,\tau)^2 dr \\
L^r_{d_0}(\tau) &= \int\limits_0^{1-\tau} V^*_{d_0}(r,\tau)^2 dr \\
W^*_{d_0}(r,\tau)^2 &= \left(W_{d_0}(r) - \tau^{-1} \int\limits_0^{\tau} W_{d_0}(r)dr \right)^2 \\
V^*_{d_0}(r,\tau)^2 &= \left(W_{d_0}(1-r) - (1-\tau)^{-1} \int\limits_{\tau}^{1} W_{d_0}(r)dr \right)^2
\end{aligned}
$$

for the de-meaned case ($z_t = 1$) *and*

$$
\begin{aligned}
L^f_{d_0}(\tau) &= \int\limits_0^{\tau} W^{**}_{d_0}(r,\tau)^2 dr \\
L^r_{d_0}(\tau) &= \int\limits_0^{1-\tau} V^{**}_{d_0}(r,\tau)^2 dr \\
W^{**}_{d_0}(r,\tau) &= W_{d_0}(r) - B_0(\tau) - rB_1(\tau) \\
V^{**}_{d_0}(r,\tau) &= -(W_{d_0}(1) - W_{d_0}(1-r)) - B^r_0(\tau) - rB^r_1(\tau) \\
B_0(\tau) &= 4\tau^{-1} \int_0^{\tau} W_{d_0}(r)dr - 6\tau^{-2} \int_0^{\tau} rW_{d_0}(r)dr \\
B_1(\tau) &= 6\tau^{-2} \int_0^{\tau} W_{d_0}(r)dr + 12\tau^{-3} \int_0^{\tau} rW_{d_0}(r)dr \\
B^r_0(\tau) &= 4(1-\tau)^{-1} \left(\int_{\tau}^{1} W_{d_0}(r)dr - (1-\tau)W_{d_0}(1) \right) \\
&\quad -6(1-\tau)^{-2} \left(\frac{(1-\tau)^2}{2} W_{d_0}(1) - \int_{\tau}^{1} W_{d_0}(r)dr \int_{\tau}^{1} rW_{d_0}(r)dr \right) \\
B^r_1(\tau) &= 6(1-\tau)^{-2} \left(\int_{\tau}^{1} W_{d_0}(r)dr - (1-\tau)W_{d_0}(1) \right) \\
&\quad +12(1-\tau)^{-3} \left(\frac{(1-\tau)^2}{2} W_{d_0}(1) - \int_{\tau}^{1} W_{d_0}(r)dr + \int_{\tau}^{1} rW_{d_0}(r)dr \right)
\end{aligned}
$$

for the de-trended case ($z_t = [1,\ t]'$).

Theorem 1 shows that the limiting distribution strongly depends on the memory parameter. This behavior of the limiting distribution leads to heavy size distortions of the original Leybourne et al. (2007) test when long-range dependencies are neglected. Therefore, we recommend to use new critical values that are provided in section 2.4. However, as the critical values vary quite substantially with d this is not very handy in practice as new critical values have to be simulated for each value of d. Therefore, we also give response curves for finite sample critical values depending on d which are easy to implement and allow a fast computation of the critical values. As the variation of the critical values with sample size is minor this is an easy procedure for practical applications, see section 2.4 for further details on this issue. After establishing the limiting distribution under the null we have to prove consistency of the test.

Theorem 2. *Let $0 \leq d_1 < 1/2$ and $1/2 < d_2 < 3/2$.*

1. *Under the alternative of a break from stationary to non-stationary long memory, this is from $I(d_1)$ to $I(d_2)$, we obtain*

$$R = O_P(T^{d_1 - d_2}).$$

2. *Under the alternative of a break from non-stationary to stationary long memory, this is from $I(d_2)$ to $I(d_1)$ we obtain*

$$R = O_P(T^{d_2 - d_1}).$$

These results imply that a consistent test against the alternative of a break from non-stationary to stationary long memory is obtained by using critical values from the upper tail of the distribution whereas using the lower tail of the distribution leads to a consistent test against the alternative of breaking from stationary to non-stationary long memory.

Remark: Although in both cases the break was assumed to be from stationary to non-stationary long memory or vice versa the test also has high power if the break is from stationary to stationary or from non-stationary to non-stationary long memory.

So far the breakpoint was assumed to be unknown. We therefore show that the breakpoint estimators given in Leybourne et al. (2007) are also consistent in the long memory setup.

Theorem 3. *Denote by τ_0 the true breakpoint. Then, for*

$$\hat{\tau} = \inf_{\tau \in \Lambda} K^f(\tau)$$

we have

$$\hat{\tau} \xrightarrow{P} \tau_0,$$

if the alternative is a break from stationary to non-stationary long memory. For a break from non-stationary to stationary long memory a consistent estimator for the breakpoint τ_0 is given by

$$\hat{\tau} = \inf_{\tau \in \Lambda} K^r(\tau).$$

As usual for long memory processes the convergence is slower than in the Leybourne et al. (2007) situation. However, this does not change the consistency result in general. Finally, we have to evaluate the behavior of the test when $X_t \sim I(d_0)$ $\forall t$ with $0 \leq d_0 < 1/2$. This is the situation where no break in persistence occurs but on the other hand our null hypothesis from Theorem 1 is wrong as the data generating process exhibits stationary long memory. We have

Theorem 4. *Let $X_t \sim I(d_0)$ $\forall t$ with $0 \leq d_0 < 1/2$. Then,*

$$R \xrightarrow{P} 1.$$

In this situation the test has a degenerated limiting distribution. As the limit theorems for non-stationary long memory processes hold for $1/2 < d_0 < 3/2$, it is reasonable to integrate the time series in this case before applying the adjusted test. For $0 \leq d_0 < 1/2$ the memory parameter of the integrated series is between 1 and 3/2. Thus, the results in Theorem 1 to Theorem 3 still hold in this situation allowing us to construct a consistent and correctly sized test. By this approach we can overcome the problem of the Leybourne et al. (2007) test to have a degenerated limiting distribution when the original series is stationary and therefore obtaining a conservative test in this situation.

Table 2.1: Empirical Size using unadjusted Critical Values

	de-meaning						de-trending					
d	1.0L	5.0L	10.0L	10.0U	5.0U	1.0U	1.0L	5.0L	10.0L	10.0U	5.0U	1.0U
0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.60	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.75	0.0	0.0	1.7	1.5	0.0	0.0	0.0	1.0	2.1	2.4	0.1	0.0
0.90	0.0	2.5	6.5	6.9	2.3	0.0	0.0	3.3	6.2	6.3	2.6	0.0
1.00	1.1	5.2	10.0	10.3	5.1	0.9	1.1	5.2	9.8	10.1	4.9	1.2

Notes: Sample size is $T = 500$, xL and xU denote the x-th lower and upper quantile of R under long-range dependencies, respectively.

Up to now, the value of d_0 has been assumed to be known. Of course, the true value of d_0 is unknown in practice and has to be estimated. However, our Monte Carlo results in the following section show that the adjusted test performs well if d_0 is estimated by a consistent estimator.

2.4 Monte Carlo Study

In this section we evaluate the finite sample behavior of the adjusted CUSUM of squares-type test. All simulations are computed in the open-source statistical programming language R, see Development Core Team R (2004). We consider the sample size $T = 500$ and use $M = 2,000$ Monte Carlo repetitions for each experiment, while simulated critical values are obtained by setting $T = 10,000$ and $M = 20,000$. The memory parameter d is treated as unknown in order to achieve realistic conditions and it is therefore estimated. We use the log-periodogram regression introduced by Geweke and Porter Hudak (1983) with a rate of frequencies of $o(T^{0.8})$ which is MSE-optimal.

First, we investigate the behavior of the Leybourne et al. (2007) test when the DGP exhibits long-range dependencies without a break in persistence, i.e. $(1-B)^d X_t = \varepsilon_t$ with

Table 2.2: Empirical Size using estimated Response Curves

	de-meaning						de-trending					
d	1.0L	5.0L	10.0L	10.0U	5.0U	1.0U	1.0L	5.0L	10.0L	10.0U	5.0U	1.0U
0.00	1.3	5.6	10.7	9.5	4.8	1.1	0.8	5.1	10.0	9.6	5.1	1.2
0.10	1.1	5.0	9.8	10.0	4.9	1.2	1.4	4.9	9.4	10.1	4.9	1.1
0.25	1.0	5.7	10.9	9.8	4.9	1.2	1.2	4.8	9.8	9.1	4.4	0.7
0.40	0.8	4.9	9.8	10.9	5.7	1.8	1.0	3.8	10.5	9.0	4.9	0.9
0.60	0.9	5.4	11.0	11.1	6.1	1.6	1.5	5.8	11.0	11.2	5.5	1.1
0.75	0.6	4.0	8.4	9.8	4.8	0.7	1.5	5.3	11.2	9.4	4.4	1.1
0.90	0.7	5.0	10.6	8.4	4.1	0.5	0.8	5.2	10.3	9.4	4.5	0.9
1.00	1.0	5.3	11.2	9.7	4.9	1.0	0.6	5.2	10.0	9.5	5.0	1.4

Notes: Sample size is $T = 500$, xL and xU denote the x-th lower and upper quantile of R under long-range dependencies, respectively.

$\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$. The long memory parameter d takes the values 0.00, 0.10, 0.25, 0.40, 0.60, 0.75, 0.90 and 1.00. This means that both, the null and the alternative hypothesis of the Leybourne et al. (2007) test are wrong except for the case of $d = 1$. As the Leybourne et al. (2007) test is known to be conservative for a process being constantly $I(0)$ we would expect a similar behavior in our setup. As we can see from Table 2.1 the originally proposed test exhibits serious size distortions in the presence of long-range dependencies resulting in a conservative test. Unsurprisingly, the empirical size is closer to the nominal significance level as d approaches one. The test is correctly sized if $d = 1$. However, even for non-stationary DGPs the size distortions are not negligible. These results underline the need for adjusted critical values taking into account the long-range dependencies.

Next, we simulate the asymptotic distribution of the R statistic dependening on d in the following 99 cases: $d = 0.51, ..., 1.49$. Due to the fact that adjusted critical values depend on d they have to be tabulated for a wide range of possible values of d. As this is rather burdensome we fit polynomial functions in d to the sequence of critical values depending on d. This response curve is given by

$$q_\alpha(d) = \sum_{i=0}^{s} \beta_i d^i \ , \tag{2.2}$$

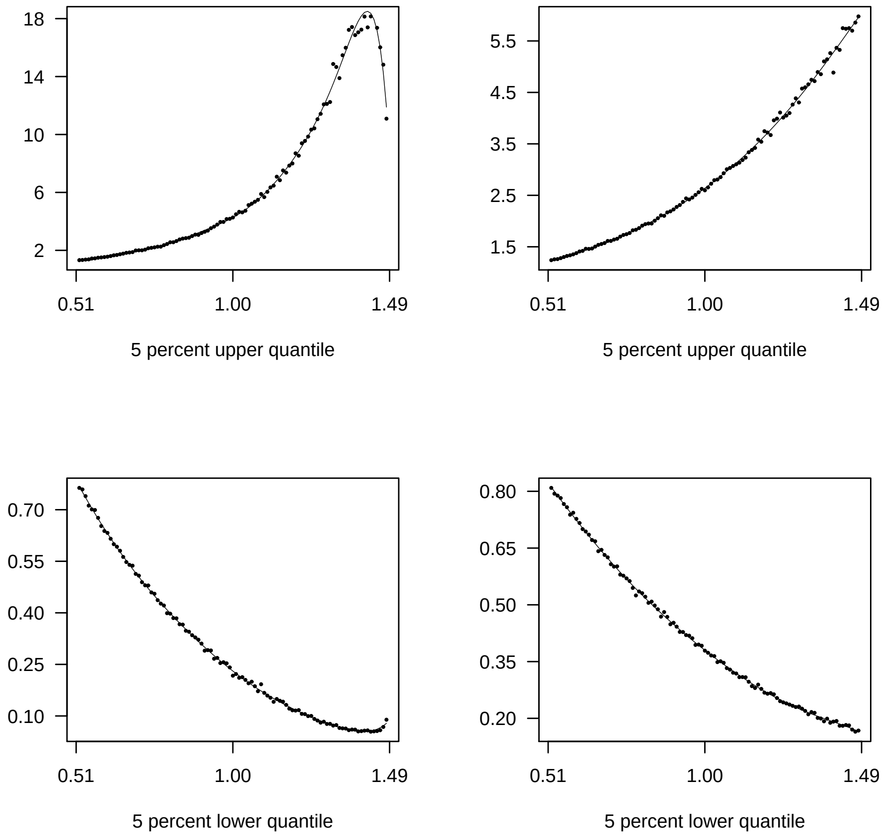


Figure 2.1: Simulated Quantiles of the R Statistic under Long Memory with fitted Response Curves for de-meaned (left) and de-trended Data (right).

where q_α denotes the α-quantile of the asymptotic distribution of R and takes values on the grid 0.51,...,1.49 consisting of 99 equally spaced points. The polynomial order s is set equal to nine. Additionally, we tried other settings, but $s = 9$ appeared to be a satisfying choice. Parameters β_i are estimated via OLS. Following the general-to-specific approach, we eliminate in each step the most insignificant power of d and re-estimate the function by OLS until no further non-rejection of the hypotheses $H_0 : \beta_i = 0$ occurs at the five percent level of significance. The final estimates are reported in Tables 2.6 and 2.7 given in Appendix B of this chapter. In Figure 2.1 we display the simulated quantiles (y-axis) depending on $d = 0.51, ..., 1.49$ (x-axis) and the fit of the response curves (solid line) for the 5% and 95% level of significance for de-meaned and de-trended data. Using the fitted response curves we can approximate critical values easily. Beside the simplicity of this approach it is reasonable in our opinion as the variation of critical values depends on the memory parameter and not on the sample size. The results in Table 2.2 that we discuss in a moment underline this argument.

Using adjusted critical values obtained from the fitted response curves we now revisit the empirical size of the test. The results are reported in Table 2.2. Note that time series with $\hat{d}_0 < 1/2$ are integrated in order to avoid a degenerated limiting distribution under the null hypothesis. Hence, we actually consider a range of values from $d = 0.60$ to 1.40. The test has satisfying properties even though there are some minor distortions for values of d in the neighborhood of 1/2. It might happen that the estimated value of d is less (greater) then 1/2 under H_0 when the true value is greater (less) than 1/2, which means that we wrongly integrate (not integrate) the time series and therefore obtain a biased test result. However, the results in Table 2.2 suggest that this is not really a serious problem.

Next we consider the power of the test based on adjusted critical values. As the test can be seen as correctly sized, there is no need for size-adjusted critical values. For all power experiments we consider three different locations of the breakpoint, at the beginning ($\tau = 0.3$), the middle ($\tau = 0.5$) and the end ($\tau = 0.7$) of the sample period. The long memory parameter takes the same values as before. The simulation results are given in Table 2.3 and 2.4 for de-meaned and de-trended data, respectively. We consider breaks from stationary to non-stationary long memory (upper left part of Tables 2.3 and 2.4) and vice versa (upper right part of Tables 2.3 and 2.4). The consistency results in Theorem 2 suggest that the test is consistent in both cases.

Furthermore, in the lower part of Tables 2.3 and 2.4 we consider breaks inside of the stationary ($0 \leq d < 1/2$) (lower left part of Tables 2.3 and 2.4) and non-stationary region ($1/2 < d < 3/2$) (lower right part of Tables 2.3 and 2.4). Note that the latter

Table 2.3: Power Experiment with de-meaned Data

d	τ			d	τ		
	0.3	0.5	0.7		0.3	0.5	0.7
$0.00 \rightarrow 0.60$	81.9	84.1	78.3	$1.00 \rightarrow 0.00$	98.7	100.0	100.0
$0.00 \rightarrow 0.75$	99.2	97.8	87.9	$1.00 \rightarrow 0.10$	99.1	100.0	100.0
$0.00 \rightarrow 0.90$	100.0	100.0	97.0	$1.00 \rightarrow 0.25$	99.5	100.0	99.1
$0.00 \rightarrow 1.00$	100.0	100.0	99.6	$1.00 \rightarrow 0.40$	99.3	99.7	95.8
$0.10 \rightarrow 0.60$	69.0	78.4	64.3	$0.90 \rightarrow 0.00$	93.4	99.9	100.0
$0.10 \rightarrow 0.75$	96.7	97.2	82.7	$0.90 \rightarrow 0.10$	95.4	99.8	99.7
$0.10 \rightarrow 0.90$	99.9	100.0	97.2	$0.90 \rightarrow 0.25$	96.1	100.0	98.1
$0.10 \rightarrow 1.00$	100.0	100.0	99.8	$0.90 \rightarrow 0.40$	96.6	98.5	89.6
$0.25 \rightarrow 0.60$	42.1	70.6	50.7	$0.75 \rightarrow 0.00$	65.5	94.5	99.0
$0.25 \rightarrow 0.75$	86.5	97.5	83.2	$0.75 \rightarrow 0.10$	69.9	95.7	97.0
$0.25 \rightarrow 0.90$	98.2	100.0	98.0	$0.75 \rightarrow 0.25$	80.9	96.7	85.8
$0.25 \rightarrow 1.00$	99.5	100.0	99.6	$0.75 \rightarrow 0.40$	84.8	88.0	59.6
$0.40 \rightarrow 0.60$	20.8	50.0	51.2	$0.60 \rightarrow 0.00$	20.2	48.6	67.1
$0.40 \rightarrow 0.75$	54.7	88.7	86.4	$0.60 \rightarrow 0.10$	25.4	55.1	61.3
$0.40 \rightarrow 0.90$	87.4	98.9	98.0	$0.60 \rightarrow 0.25$	37.5	64.3	40.9
$0.40 \rightarrow 1.00$	95.8	99.9	99.6	$0.60 \rightarrow 0.40$	49.4	47.8	19.2
$0.00 \rightarrow 0.10$	82.7	84.8	82.5	$1.00 \rightarrow 0.60$	81.9	86.4	68.5
$0.00 \rightarrow 0.25$	77.4	82.0	81.0	$1.00 \rightarrow 0.75$	45.1	45.1	36.8
$0.00 \rightarrow 0.40$	64.6	79.6	77.0	$1.00 \rightarrow 0.90$	13.9	14.5	12.4
$0.10 \rightarrow 0.25$	59.5	68.1	67.3	$0.90 \rightarrow 0.60$	68.5	65.9	46.3
$0.10 \rightarrow 0.40$	48.5	62.9	61.9	$0.90 \rightarrow 0.75$	24.7	25.0	15.8
$0.25 \rightarrow 0.40$	21.7	33.1	32.3	$0.75 \rightarrow 0.60$	36.4	31.4	16.2

Table 2.4: Power Experiment with de-trended Data

d	τ			d	τ		
	0.3	0.5	0.7		0.3	0.5	0.7
$0.00 \rightarrow 0.60$	45.1	84.7	94.2	$1.00 \rightarrow 0.00$	100.0	99.9	97.2
$0.00 \rightarrow 0.75$	64.1	95.8	99.5	$1.00 \rightarrow 0.10$	100.0	99.9	96.6
$0.00 \rightarrow 0.90$	88.6	99.7	99.9	$1.00 \rightarrow 0.25$	99.9	99.6	95.2
$0.00 \rightarrow 1.00$	96.5	99.9	100.0	$1.00 \rightarrow 0.40$	99.0	98.8	89.3
$0.10 \rightarrow 0.60$	40.8	76.6	87.1	$0.90 \rightarrow 0.00$	99.9	99.7	90.8
$0.10 \rightarrow 0.75$	61.1	93.7	97.6	$0.90 \rightarrow 0.10$	99.6	99.8	90.0
$0.10 \rightarrow 0.90$	89.3	99.6	99.7	$0.90 \rightarrow 0.25$	98.5	98.6	85.4
$0.10 \rightarrow 1.00$	97.1	99.9	99.9	$0.90 \rightarrow 0.40$	97.1	95.6	73.7
$0.25 \rightarrow 0.60$	33.6	57.4	64.9	$0.75 \rightarrow 0.00$	99.1	96.1	64.8
$0.25 \rightarrow 0.75$	53.7	87.4	92.0	$0.75 \rightarrow 0.10$	96.5	94.6	63.1
$0.25 \rightarrow 0.90$	85.5	98.4	99.0	$0.75 \rightarrow 0.25$	91.0	88.6	57.2
$0.25 \rightarrow 1.00$	95.0	99.7	99.8	$0.75 \rightarrow 0.40$	80.3	71.3	42.8
$0.40 \rightarrow 0.60$	19.6	33.1	41.7	$0.60 \rightarrow 0.00$	94.1	84.3	46.3
$0.40 \rightarrow 0.75$	39.1	71.7	77.9	$0.60 \rightarrow 0.10$	85.9	75.9	42.6
$0.40 \rightarrow 0.90$	72.5	94.4	96.4	$0.60 \rightarrow 0.25$	65.4	58.7	32.7
$0.40 \rightarrow 1.00$	89.3	98.5	99.5	$0.60 \rightarrow 0.40$	42.1	34.5	19.0
$0.00 \rightarrow 0.10$	7.3	13.5	21.8	$1.00 \rightarrow 0.60$	85.0	85.3	63.6
$0.00 \rightarrow 0.25$	27.5	39.4	49.5	$1.00 \rightarrow 0.75$	48.0	47.1	34.4
$0.00 \rightarrow 0.40$	52.4	69.3	75.8	$1.00 \rightarrow 0.90$	14.5	15.7	12.3
$0.10 \rightarrow 0.25$	12.8	20.5	26.0	$0.90 \rightarrow 0.60$	65.5	64.7	39.8
$0.10 \rightarrow 0.40$	33.6	46.6	53.5	$0.90 \rightarrow 0.75$	23.4	24.0	17.6
$0.25 \rightarrow 0.40$	13.4	16.3	25.7	$0.75 \rightarrow 0.60$	29.3	26.6	16.6

Table 2.5: Small Sample Performance of Breakpoint Estimators

		de-meaning			de-trending		
d	τ	0.3	0.5	0.7	0.3	0.5	0.7
$U[0,0.4] \rightarrow U[0.6,1]$	$\hat{\tau}^f$	0.378	0.546	0.725	0.442	0.557	0.730
	se($\hat{\tau}^f$)	0.139	0.076	0.035	0.183	0.087	0.038
$U[0.6,1] \rightarrow U[0,0.4]$	$\hat{\tau}^r$	0.275	0.453	0.610	0.271	0.440	0.568
	se($\hat{\tau}^r$)	0.034	0.075	0.148	0.036	0.087	0.178
$U[0,0.2] \rightarrow U[0.8,1]$	$\hat{\tau}^f$	0.327	0.521	0.717	0.340	0.524	0.717
	se($\hat{\tau}^f$)	0.058	0.037	0.026	0.078	0.043	0.026
$U[0.8,1] \rightarrow U[0,0.2]$	$\hat{\tau}^r$	0.281	0.477	0.672	0.279	0.473	0.660
	se($\hat{\tau}^r$)	0.025	0.039	0.055	0.027	0.044	0.076

Notes: Sample size is $T = 500$, $U[i,j]$ denotes the uniform distribution with lower and upper bound i and j, respectively; $\hat{\tau}^f$ and $\hat{\tau}^r$ denote the break point estimator based on the forward (f) and reversed series (r), respectively; se$(\cdot)$ is the standard error.

experiments are not covered by any of our theorems but that they might be relevant in empirical applications.

Overall, the power results of the adjusted test using de-meaned data (Table 2.3) are good and confirm the consistency result. For quite extreme breaks, e.g. 0.00 to 0.90, the power is almost hundred percent. Unsurprisingly, the power decreases for less extreme breaks, e.g. 0.00 to 0.60. For a given value of d_2, the power decreases with increasing d_1 (left part). For a given value of d_2, the power decreases with decreasing d_1 (right part). At first sight, it might not be intuitive that the power for breaks of equal distance, e.g. 0.25 to 0.75 and 0.40 to 0.90, is not the same. This is an artefact of the adjusted test introduced by integrating the time series if the estimated value of d is located in the stationary region. In addition, the fact that the breakpoint influences the estimate of d under H_0 further complicates the interpretation. However, the test is able to detect switches of the long memory parameter within the stationary and non-stationary region. The main conclusions are not changing when looking at the results for de-trended data.

After evaluating the power we consider the small sample performance of the simple breakpoint estimator, see Theorem 3. We use the same three different break points as before, i.e. $\tau = 0.3, 0.5, 0.7$. The memory parameter switches from the stationary to the non-stationary region and vice versa. For both regions we draw the memory parameter from a uniform distribution in order to cover a wide range of possible values in a small number of experiments. The upper and lower bounds are set equal to $[0, 0.4]$ and $[0.6, 1]$ for the stationary and the non-stationary region, respectively. In a more restrictive setting we set them equal to $[0, 0.2]$ and $[0.8, 1]$. Results for a sample size of five hundred observations are reported in Table 2.5. The overall impression of the breakpoint estimator's performance is satisfying. Noteworthy, we observe that the breakpoint estimator performs worse in the case of de-trended data which is due to an additional nuisance parameter that has to be estimated, c.f. Leybourne et al. (2007). We further note that $\hat{\tau}^f$ and $\hat{\tau}^r$ perform better in the more restrictive setting because the break point becomes easier to detect.

2.5 Empirical Illustration

In this section we consider an empirical application of the adjusted test to study whether there is any change in persistence in US inflation. Hassler and Wolters (1995) reported that such time series exhibit long-range dependencies. In addition, there is a small but growing literature dealing with the persistence of US inflation. Main questions in this literature are the measurement of persistence and potential structural breaks in it. In a recent article, Pivetta and Reis (2007) come to the conclusion that persistence of US inflation is approximately constant over time. The authors argue that their conclusion is in line with Stock and Watson (2003) as well as O'Reilly and Whelan (2005). However, Kang et al. (2006) provide evidence for a decline in persistence at the very end of the seventies. Nonetheless, to the best of our knowledge, previous studies are not concerned with long-range dependencies and there is no previous study dealing with a structural break in the long memory parameter regarding inflation time series. Therefore, we try to add some new evidence by applying our long memory adjusted test.

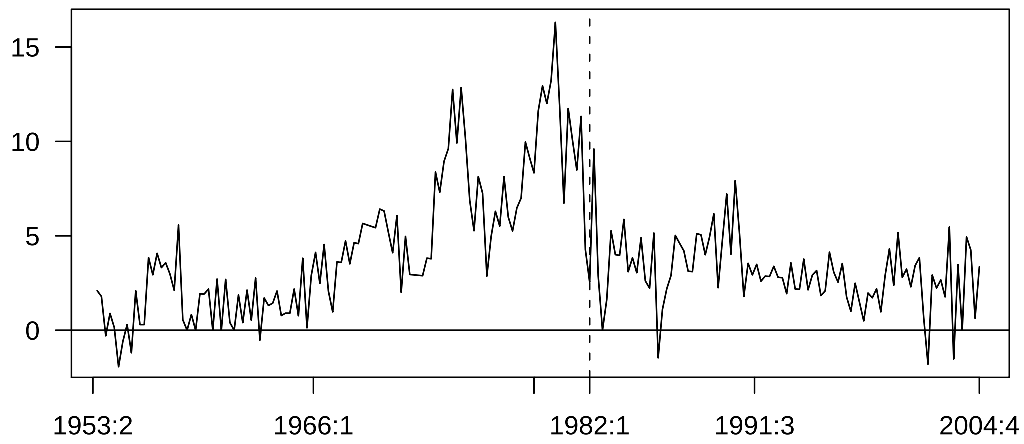


Figure 2.2: Time Series Plot of US Inflation.

We use the quarterly CPI data from Lanne (2006) and transform it to annualized inflation by computing $y_t = 400 \ln(\text{CPI}_t/\text{CPI}_{t-1})$. The CPI data spans from 1953:1 to 2004:4 implying 207 observations. The time series plot is depicted in Figure 2.2. The graph suggests a decline of persistence in the second half of the sample which might be a result of Volcker's policy to pull inflation down from its high level in the seventies. The vertical line at 1982:1 shows the estimated break point of our test which will be discussed below.

Under the null hypothesis we obtain an estimated value of d_0 via the GPH approach with MSE-optimal rate of frequencies that equals 0.617 indicating non-stationary long-memory. We apply the adjusted test without integrating the time series under consideration since the asymptotic distribution of the test statistic is not degenerated as long as $d_0 > 1/2$ holds. Furthermore, we de-mean the data in a first step, since a clear linear trend is not obvious. Testing the null hypothesis of constant memory against decreasing memory gives a test statistic of 1.801 which is significant at the ten and five percent level of significance. Note that we make use of our estimated response curves to approximate the relevant critical values. The test result suggests that there is a decline in the persistence of US inflation. Interestingly, the estimated break point is 1982:1 which is nine quarters after the begin of Volcker's chairmanship at the Federal Reserve. When estimating the long memory parameter d before and after this breakpoint we get 0.862 and 0.246 which can be viewed as a sharp decline in persistence. Finally, we test the null hypothesis of constant memory for the time period after 1982:1. Note that we have to integrate the time series once, since the asymptotic distribution is degenerated for $0 \leq d_0 < 1/2$. The test statistic is now 1.667 and insignificant at conventional levels. Although this result is based on only 91 observations, it suggests that there is no additional break after 1982:1.

2.6 Conclusion

In this chapter we present a modification of a test proposed by Leybourne et al. (2007) that allows for long memory dynamics. In particular, the test is constructed for the null hypothesis that there is no change in the long-memory parameter d against the alternative that it breaks from a stationary value ($0 \leq d < 1/2$) to a non-stationary one ($1/2 < d < 3/2$) or vice versa. We derive several asymptotic properties of the test statistic under long-range dependent DGPs and show that the asymptotic distribution depends on d. Therefore, we propose response curves based on estimates for d to obtain the relevant critical value easily and show by means of a Monte Carlo study that this approach works well. Furthermore, the power of the test is good and a simple breakpoint estimator has satisfying properties. Finally, we apply the test to US inflation data and find a break from non-stationary to stationary long-memory in the early eighties.

2.7 Appendix A

Proof of Theorem 1: For the proof of the theorem let us consider the de-meaned case first. The test statistic was defined by

$$R = \frac{\inf_{\tau \in \Lambda} K^f(\tau)}{\inf_{\tau \in \Lambda} K^r(\tau)},$$

with

$$K^f(\tau) = [\tau T]^{-2} \sum_{t=1}^{[\tau T]} \hat{v}_{t,\tau}^2$$

and

$$K^r(\tau) = (T - [\tau T])^{-2} \sum_{t=1}^{T-[\tau T]} \tilde{v}_{t,\tau}^2.$$

For the nominator we have

$$T^{-d_0-\frac{1}{2}} \hat{v}_{t,\tau} = T^{-d_0-\frac{1}{2}} x_t - T^{-d_0-\frac{1}{2}} \bar{x}(\tau).$$

We have

$$T^{-d_0-\frac{1}{2}} x_{[rt]} \Rightarrow W_{d_0}(r)$$

with W_{d_0} denoting fractional Brownian motion with parameter d_0. Furthermore we have

$$\begin{aligned} T^{-d_0-\frac{1}{2}} \bar{x}(\tau) &= T^{-d_0-\frac{1}{2}} [\tau T]^{-1} \sum_{t=1}^{[\tau T]} x_t \\ &= T^{-d_0-\frac{3}{2}} \tau^{-1} \sum_{t=1}^{[\tau T]} x_t \\ &\Rightarrow \tau^{-1} \int_0^\tau W_{d_0}(r) dr. \end{aligned}$$

Application of the continuous mapping theorem gives

$$\begin{aligned} T^{-d_0-\frac{1}{2}} \hat{v}_{[rt]} &\Rightarrow W_{d_0}(r) - \tau^{-1} \int_0^\tau W_{d_0}(r) dr \\ &=: W_{d_0}^*(r,\tau) \end{aligned}$$

and thus for the nominator

$$\begin{aligned} T^{-2d_0} K^f(\tau) &= \tau^{-2} \int_0^\tau (T^{-d_0-\frac{1}{2}} \hat{v}_{[rT]})^2 dr \\ &\Rightarrow \tau^{-2} \int_0^\tau W_{d_0}^*(r,\tau)^2 dr. \end{aligned}$$

Similarly we obtain for the denominator

$$\begin{aligned} T^{-d_0-\frac{1}{2}}\tilde{v}_{[rT]} &\Rightarrow W_{d_0}(1) - W_{d_0}(1-r) + (1-\tau)^{-1}\int_0^{1-\tau}(W_{d_0}(1) - W_{d_0}(1-r)dr) \\ &= W_{d_0}(1-r) - (1-\tau)^{-1}\int_\tau^1 W_{d_0}(r)dr \\ &=: V^*_{d_0}(r,\tau). \end{aligned}$$

Again using the continuous mapping theorem we obtain for the denominator

$$\begin{aligned} T^{-2d_0}K^r(\tau) &= (1-\tau)^{-2}\int_0^{1-\tau}(T^{-d_0-\frac{1}{2}}\tilde{v}_{[rT]})^2 dr \\ &\Rightarrow (1-\tau)^{-2}\int_0^{1-\tau} V^*_{d_0}(r,\tau)^2 dr. \end{aligned}$$

Combining the result for the nominator and the denominator gives the result.

The result for the de-meaned and de-trended case is obtained by applying standard results for linear regression with long-memory errors. We consider the forward statistic

$$K^f(\tau) = [\tau T]^{-2}\sum_{t=1}^{[\tau T]}\hat{v}_t^2,$$

where $\hat{v}_t = x_t - \hat{\alpha} - \hat{\beta}t$ are the residuals from the OLS regression of x_t on the vector $z_t = [1,t]'$, $t = 1,\ldots,[\tau T]$. It is well known that $T^{-d_0-1/2}(\hat{\alpha}-\alpha) \Rightarrow B_0(\tau)$ and $T^{-d_0-1/2}(\hat{\beta}-\beta) \Rightarrow B_1(\tau)$ and $B_0(\tau)$ and $B_1(\tau)$ given as in the Theorem. Therefore, we obtain

$$\begin{aligned} T^{-d_0-1/2}\hat{v}_{[\tau T]} &= T^{-d_0-1/2}v_{[\tau T]} - T^{-d_0-1/2}(\hat{\alpha}-\alpha) - rT^{-d_0-1/2}(\hat{\beta}-\beta) \\ &\Rightarrow W_{d_0}(r) - B_0(\tau) - rB_1(\tau) \\ &\equiv W^{**}_{d_0}(r,\tau). \end{aligned}$$

As the forward statistic is a continuous functional of $T^{-d_0-1/2}\hat{v}_{[\tau T]}$ we obtain using the CMT

$$\begin{aligned} T^{-2d_0}K^f(\tau) &= \tau^{-2}\int_0^\tau (T^{-d_0-1/2}\hat{v}_{[\tau T]})^2 dr \\ &\Rightarrow \tau^{-2}\int_0^\tau W^{**}_{d_0}(r,\tau)^2 dr \\ &\equiv L^f(\tau). \end{aligned}$$

The proof for the reverse statistic is analogous and therefore omitted. For the remainder of this Appendix we omit proofs for the de-meaned and the de-trended case for the brevity of notation as they are straightforward. □

Proof of Theorem 2: First we prove the first part of the theorem. This is we assume a breakpoint that the DGP breaks from a stationary to a non-stationary long-memory process. Let us first consider the situation of $\tau \leq \tau_0$, where τ_0 denotes the true breakpoint. This means that $X_t \sim I(d_1)$ with $0 \leq d_1 < 1/2$. Have in mind that the standardization of the test statistic is obtained from $H_0 : X_t \sim I(d_0)$ with $d_1 \neq d_0$. In the stationary part we have $d_0 \geq d_1$. In this situation we obtain:

$$\begin{aligned} T^{-2d_0+1} K^f_{d_0}(\tau) &= \tau^{-1} T^{-2d_0} [\tau T]^{-1} \sum_{t=1}^{[\tau T]} v_t^2 \\ &\xrightarrow{P} \tau^{-1} O(T^{d_1 - d_0}). \end{aligned}$$

In the case of $d_1 = d_0$ the upper expression converges to $\tau^{-1}\gamma_0$ with γ_0 denoting the variance of X_t. For $d_1 < d_0$, wich is the relevant case in practise, this expression tends to zero with a rate depending on the difference of the true d_0 before the break and the hypothetic memory parameter.

We next consider the situation of $\tau > \tau_0$ where we split $K^f_d(\tau)$ up in its stationary and its non-stationary part. Have in mind that the true DGP is of order $1/2 < d_2 < 3/2$ after the break with $d_2 > d_0$ in the non-stationary part.

$$\begin{aligned} T^{-2d_0} K^f_{d_0}(\tau) &= \tau^{-2} T^{2-2d_0} \sum_{t=1}^{[\tau T]} x_t^2 - \tau^{-3} \left(T^{3/2-d_0} \sum_{t=1}^{[\tau T]} y_t \right)^2 \\ &= \tau^{-2} \left(T^{2-2d_0} \sum_{t=1}^{[\tau_0 T]} x_t^2 + T^{2-2d_0} \sum_{t=[\tau_0 T]+1}^{[\tau T]} x_t^2 \right) \\ &\quad -\tau^{-3} \left(T^{3/2-d_0} \sum_{t=1}^{[\tau_0 T]} y_t + T^{3/2-d_0} \sum_{t=[\tau_0 T]+1}^{[\tau T]} y_t \right)^2 \\ &= \tau^{-2} T^{2-2d_0} \sum_{t=[\tau_0 T]+1}^{[\tau T]} x_t^2 - \tau^{-3} \left(T^{3/2-d_0} \sum_{t=[\tau_0 T]+1}^{[\tau T]} y_t \right)^2 + o_P(1) \\ &\xrightarrow{P} O_P(T^{d_2 - d_0}). \end{aligned}$$

From these considerations we see that the limit of $T^{-2d_0+1}K^f_{d_0}(\tau)$ is given by $\tau^{-1}O_P(T^{d_1-d_0})$ $+ \infty 1_{(\tau>\tau_0)}$ which is obviously minimized by τ_0. Thus, we have

$$T^{-2d_0+1}\inf_{\tau\in\Lambda} K^f_{d_0}(\tau) \xrightarrow{P} O_P(T^{d_1-d_0}).$$

For the reversed series we obtain by similar arguments for $\tau \leq \tau_0$:

$$T^{-2d_0}K^r_{d_0}(\tau) \xrightarrow{P} O_P(T^{d_2-d_0}).$$

This gives us

$$\inf_{\tau\in\Lambda} K^r_{d_0}(\tau) = O_P(T^{d_2+d_0-1})$$

which gives us the first result of the theorem. The result in point 2 of Theorem 2 is obtained by similar arguments as above. □

Proof of Theorem 3: Let us assume a break from stationary long memory to non-stationary long memory, that is $0 \leq d_1 < 1/2$ and $1/2 < d_2 < 3/2$. The hypothetical memory parameter is denoted by d_0 with $d_1 \leq d_0 \leq d_2$. ¿From Theorem 2 we know that the limit of $T^{-2d_0+1}K^f_{d_0}(\tau)$ is given by $O_P(T^{d_1-d_0})1_{(\tau\leq\tau_0)} + \infty 1_{(\tau>\tau_0)}$ which is obviously minimized by τ_0. The result follows now by similar arguments as in Leybourne et al. (2007). The proof for the second part of the theorem, that is the break from non-stationary to stationary long memory, is analogous and therefore omitted here. □

Proof of Theorem 4: Because of the symmetry of Λ around 0.5 we have

$$\begin{aligned}
\inf_{\tau\in\Lambda} T^{-2d_0}K^f_{d_0}(\tau) &\xrightarrow{P} \inf_{\tau\in\Lambda} \tau^{-1}\gamma_0 \\
&= \lambda_u^{-1}\gamma_0 \\
\inf_{\tau\in\Lambda} T^{-2d_0}K^r_{d_0}(\tau) &\xrightarrow{P} \inf_{\tau\in\Lambda}(1-\tau^{-1})\gamma_0 \\
&= (1-\lambda_l^{-1})\gamma_0 \\
&= \lambda_u^{-1}\gamma_0,
\end{aligned}$$

where λ_u and λ_l denote the upper and the lower bound of the interval Λ respectively. This proves the theorem. □

2.8 Appendix B

Table 2.6: Estimated Response Curves for de-meaned Data

Quantile	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
1.0L	1.063	0	0	0	-41.002	133.627	-183.98	131.206	-47.89	7.102
5.0L	1.601	0	-7.486	9.449	0	0	-17.596	25.299	-13.724	2.688
10.0L	-221.524	2316.11	-10522.512	27414.943	-45191.318	48907.541	-34769.527	15666.998	-4062.561	462.173
10.0U	5145.518	-54469.126	252323.451	-671384.183	1131196.84	-1252080.53	910897.739	-420239.255	111628.329	-13015.697
5.0U	10493.76	-110784.01	511682.48	-1357262	2279365.93	-2514370.73	1822761.11	-837851.29	221721.78	-25752.77
1.0U	-1174.527	0	58540.259	-312952.617	792898.52	-1170633.31	1062803.45	-586254.848	180679.152	-23898.266

Notes: xL and xU denote the x-th lower and upper quantile of R. OLS estimates for β_i $(i = 0, 1, ..., 9)$ in (2.2) are reported in columns; $\beta_i = 0$ means that the parameter is set equal to zero.

Table 2.7: Estimated Response Curves for de-trended Data

Quantile	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
1.0L	1.051	0	0	-4.815	0	18.496	-25.406	13.556	-2.63	0
5.0L	1.151	0	0	-9.281	21.702	-21.366	9.999	-1.824	0	0
10.0L	-0.455	0	53.424	-234.177	459.766	-499.311	310.551	-103.809	14.485	0
10.0U	1.054	0	0	0	3.328	-3.117	0.868	0	0	0
5.0U	1.008	0	0	0	8.274	-13.18	8.509	-1.971	0	0
1.0U	1.187	0	0	0	6.272	-5.03	1.557	0	0	0

Notes: See Table 2.6.

Chapter 3

Rational Bubbles and changing Degree of Fractional Integration

Published as Hannover Leibniz University Discussion Paper No. 394, March 2008 (ISSN 0949-9962).

3.1 Introduction

In this chapter we provide evidence for a rational bubble in S&P 500 stock prices by applying a test for changing persistence under fractional integration. Koustas and Serletis (2005) find strong evidence for the existence of long memory in the S&P 500 log dividend yield and their results support the hypothesis of no rational bubble. However, the authors did not account for a potential change in the fractional degree of integration. We apply a suitable test proposed by Sibbertsen and Kruse (2007) and find a significant break in the memory of the S&P 500 log dividend yield that is located at November, 1955. This breakpoint is also found by Sollis (2006) who apply tests for a change in persistence that are designed for the $I(0)/I(1)$ framework. We find strong evidence for stationary long memory before the break in 1955 and a unit root afterwards. These results confirm on one hand the previous result of fractional integration in this time series and on the other hand they are in line with other empirical studies that found evidence for a rational bubble in it.

Section 3.2 reviews the test for changing memory briefly, while the empirical results are reported in section 3.3 and section 3.4 concludes.

3.2 Testing for changing Memory

We assume that the data generating process follows an $I(d)$ process as proposed by Granger and Joyeux (1980):

$$(1-L)^d y_t = \varepsilon_t,$$

where ε_t are i.i.d. random variables with mean zero and variance σ^2 and L denotes the lag operator ($L^k y_t \equiv y_{t-k}$). This process is said to be fractionally integrated of order d. The test proposed by Sibbertsen and Kruse (2007) considers the following pair of hypotheses,

$$\begin{aligned} H_0 &: \; d = d_0 \;\text{ for all } t \\ H_1 &: \; d = d_1 \;\text{ for }\; t = 1, \ldots, [\tau T] \\ & \quad\; d = d_2 \;\text{ for }\; t = [\tau T]+1, \ldots, T \end{aligned}$$

where $[x]$ denotes the biggest integer smaller than x. The differencing parameter is restricted to $0 \leq d_0 < 3/2$ under H_0, while $0 \leq d_1 < 1/2$ and $1/2 < d_2 < 3/2$. Note that, d_1 and d_2 can be interchanged. This means that we test the null hypothesis of constant memory against a change from stationary ($0 \leq d_1 < 1/2$) to non-stationary ($1/2 < d_2 < 3/2$) long memory at $[\tau T]$ and vice versa. The test statistic is given by

$$R = \frac{\inf_{\tau \in \Lambda} K^f(\tau)}{\inf_{\tau \in \Lambda} K^r(\tau)},$$

where $K^f(\tau)$ and $K^r(\tau)$ are CUSUM of squares-based statistics depending on the forward and reversed residuals of the data generating process as given below. The relative breakpoint $\tau \in \Lambda \subset (0,1)$ is assumed to be unknown and a simple estimator is given below. In detail, the forward and reverse CUSUM of squares-based statistics are defined by

$$K^f(\tau) = [\tau T]^{-2} \sum_{t=1}^{[\tau T]} \hat{v}_{t,\tau}^2$$

and

$$K^r(\tau) = (T - [\tau T])^{-2} \sum_{t=1}^{T-[\tau T]} \tilde{v}_{t,\tau}^2.$$

Here, $\hat{v}_{t,\tau}$ are the residuals from the OLS regression of y_t on a constant based on the observations up to $[\tau T]$. This is

$$\hat{v}_{t,\tau} = y_t - \bar{y}(\tau)$$

with $\bar{y}(\tau) = [\tau T]^{-1} \sum_{t=1}^{[\tau T]} y_t$. Similarly $\tilde{v}_{t,\tau}$ is defined for the reversed series $z_t \equiv y_{T-t+1}$. Thus, it is given by

$$\tilde{v}_{t,\tau} = z_t - \bar{z}(1-\tau)$$

with $\bar{z}(1-\tau) = (T - [\tau T])^{-1} \sum_{t=1}^{T-[\tau T]} z_t$. Since the limiting distribution of R depends on the memory parameter under the null hypothesis d_0, Sibbertsen and Kruse (2007) provide response curves that allow an easy computation of relevant critical values. Note that, when testing against a change from stationary to non-stationary memory the left tail of the distribution is relevant and vice versa. Furthermore, the authors prove consistency of the simple breakpoint estimator that is given by

$$\hat{\tau} = \inf_{\tau \in \Lambda} K^f(\tau).$$

3.3 Empirical Evidence

The used monthly data set can be downloaded from Robert Shiller's web site[1]. The sample spans from January, 1871 to December, 2007 implying 1644 observations. The time series is depicted in Figure 3.1. The graph shows a clear change in the behavior in the last third of the sample.

In a first step of our analysis, we estimate the long memory parameter by applying the log periodogram regression method proposed by Geweke and Porter-Hudak (1983). This estimator is based on an approximation of the spectral density near the origin. A crucial issue is the choice of number of frequencies m that are used to perform the log periodogram regression. Hurvich et al. (1998) show that $m = o(T^{4/5})$ is MSE-optimal. On the other hand Geweke and Porter-Hudak suggest to use $m = o(T^{1/2})$ which means that higher frequencies are disregarded which implies that the estimator is less efficient. On

[1]http://cowles.econ.yale.edu/faculty/shiller.htm/

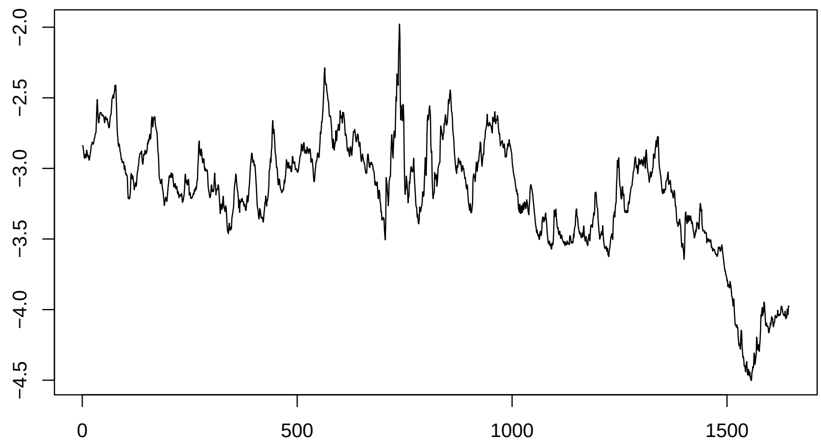


Figure 3.1: S&P 500 Log Dividend Yield (January, 1871 to December, 2007).

the other hand, if the true DGP contains short-term dependencies which are usually represented by an ARMA(p, q) process, the GPH estimator based on $m = o(T^{1/2})$ is less biased. Hence, there is a tradeoff between bias and efficiency.

Davidson and Sibbertsen (2007) recently proposed a Hausman-type test for the bias in log-periodogram regressions that compares two GPH estimators using a different number of frequencies. Under the null hypothesis short-term dependencies are negligible and therefore a higher number of frequencies, $m = o(T^{4/5})$, can be used without running the risk of a bias. Under the alternative the authors suggest to use a lower number of frequencies, $m = o(T^{1/2})$. An application of this test leads to a rejection at the nominal five percent level of significance (p-value = 0.030). The estimate of d_0 using $m = T^{1/2}$ is 0.82 which indicates a non-stationary long memory time series. Alternatively to the log periodogram regression approach we estimate an ARFIMA$(p, d, 0)$ model in order to account explicitly for short-term correlations represented by a finite AR component. MA components are omitted for simplicity. The model can be written as

$$(1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p)(1 - L)^d y_t = \mu + \varepsilon_t \ ,$$

where α_i denote the AR-parameter corresponding to lag i and μ is a constant. All parameters of this model are estimated jointly via non-linear least squares which is often referred to as the conditional sum-of-squares (CSS) estimator that has been suggested by Beran (1995) and further studied in Chung and Baillie (1993) and Doornik and Ooms (2004). In contrast to exact maximum likelihood or modified profile likelihood estimation, the NLS estimator is also applicable for non-stationary ARFIMA models ($0.5 < d \leq 1$). The optimal autoregressive lag length is chosen via AIC with a lower bound of zero and an upper bound of $p_4 = [4(T/100)^{1/4}] = 7$, cf. Schwert (1989).

Detailed NLS estimation results can be found in Table 3.1. Compared to the GPH estimate we obtain a slightly lower value of 0.61. A simple t-test of the null hypothesis of short memory $H_0 : d_0 = 0$ has to be strongly rejected. The Ljung-Box statistic Q with 12 lags is not significant which suggests that there is no remaining autocorrelation up to lag 12 left in the residuals.

The interval of potential breakpoints is set as $\Lambda = [0.2, 0.8]$ which is a common choice in the literature. The test statistic R equals 0.35. Based on $\hat{d}_0 = 0.82$, critical values that are computed via response curves equal 0.47, 0.38 and 0.24 for the nominal 10, 5 and 1 percent level of significance, respectively. We have to reject the null hypothesis of constant memory in favor of the alternative that the memory increases for small values of R. Thus, we find evidence for changing memory at the five percent level. When using the ARFIMA model based estimate of d_0 the critical values are 0.70, 0.61 and 0.48, respectively. Thus, H_0 has to be rejected even at the one percent level of significance. We therefore conclude, that there might be a change in d from d_1 to d_2.

The estimated breakpoint is at observation 1019 which corresponds to November, 1955. Sollis (2006) finds a very similar breakpoint by applying the Leybourne et al. (2003) test for a unit root against a change from $I(0)$ to $I(1)$. The GPH estimate of the memory parameter before the break (based on $T_1 = 1019$ observations) is $\hat{d}_1 = 0.37$ and signifi-

Table 3.1: NLS Estimation Results for ARFIMA(p,d,0) Models

	$t = 1 - 1644$		$t = 1 - 1019$		$t = 1020 - 1644$	
μ	-3.079	(0.000)	-2.949	(0.000)		
d	0.612	(0.000)	0.327	(0.006)	0.949	(0.000)
α_1	0.712	(0.000)	1.011	(0.000)	0.324	(0.003)
α_2	-0.071	(0.162)	-0.191	(0.099)	-0.057	(0.227)
α_3	0.012	(0.787)	-0.021	(0.754)	0.050	(0.343)
α_4	0.061	(0.314)	0.114	(0.068)	-0.011	(0.806)
α_5	0.080	(0.048)			0.122	(0.015)
α_6					-0.086	(0.057)
$Q(12)$	4.717	(0.967)	7.769	(0.803)	2.938	(0.996)

Notes: P-values are reported in brackets beside the corresponding estimate or test statistic.

cantly different from zero (p-value = 0.000). This result suggests that the S&P 500 log dividend yield is fractionally integrated before November, 1955. Considering the estimated ARFIMA model for the first sub-sample, we find further evidence for long-range dependence since the estimate $\hat{d}_1 = 0.327$ is highly significant (p-value = 0.006). After the break the GPH estimate increases to $\hat{d}_2 = 1.09$ which is close to unity but higher than one suggesting a potential unit root. This estimate is based on $T_2 = 625$ observations. Again, the ARFIMA model based estimate ($\hat{d}_2 = 0.949$) is lower but even closer to unity.

In order to carry out a formal and suitable test of the unit root hypothesis against long memory we apply the fractional Dickey-Fuller test proposed by Dolado et al. (2002). Their procedure is based on the test regression

$$\Delta^{\delta_0} y_t = \phi \Delta^{\delta_1} y_{t-1} + \sum_{i=1}^{p} \lambda_i \Delta y_{t-i} + \varepsilon_t$$

where $\delta_0 = 1$ in our application, δ_1 is unknown and has to be estimated from the data. Note, that the estimator for δ_1 has to be $T^{1/2}$-consistent, we therefore employ the parametric NLS estimator proposed by Beran (1995). Regarding the test regression the relevant pair of hypotheses is $H_0 : \phi = 0$ versus $H_1 : \phi < 0$. Dolado et al. (2002) prove that the limiting distribution of the t-statistic for H_0 is standard normal if $0.5 \leq \delta_1 < 1$ which is the relevant case in our application. For further details, the reader is referred to Dolado et. al (2002). As before, the maximum lag length is set equal to $p_4 = [4(T_2/100)^{1/4}] = 6$. The optimal length is chosen with the Schwarz information criterion and equals zero.The estimated test regression without lags of Δy_t is given by

$$\Delta y_t = 0.245\Delta^{0.949} y_{t-1} + \hat{\varepsilon}_t \ .$$

Since the test statistic $t_\phi = 6.593$ is not significant at conventional levels, we are not able to reject the null hypothesis of a unit root in the second sub-sample which hints at a rationale bubble because the no-bubbles restriction is not fulfilled in this case.

3.4 Conclusions

This chapter provides evidence that the time series properties of the S&P 500 log dividend yield are changing over time. By applying recent tests we find that the time series is stationary and fractionally integrated before November, 1955 and that it exhibits a unit root afterwards. The presence of a unit root in the second sub-sample suggests a rationale bubble in the S&P 500 stock price.

Chapter 4

Unit Root Testing against ESTAR with modified Statistics

Published as Hannover Leibniz University Discussion Paper No. 398, April 2008 (ISSN 0949-9962) and as a revision in Statistical Papers (2009), forthcoming, DOI: 10.1007/s00362-009-0204-1.

4.1 Introduction

Non-linear time series models like the smooth transition autoregressive (STAR) model, see Teräsvirta (1994), have become very popular in the last years. In this chapter we are mainly concerned with the exponential STAR (ESTAR) model and develop a new test for the unit root hypothesis against a globally stationary ESTAR model. In particular, we focus on a prominent and widely applied specification of this model that allows for a unit root regime and two symmetric mean-reverting regimes. The time series process, say y_t, behaves like a random walk if y_{t-1} was close to some location parameter c and it is mean-reverting if y_{t-1} departs from c. In the exponential smooth transition model the degree of mean-reversion depends on the squared difference between y_{t-1} and c.

When modeling real exchange rates for example, the economic intuition behind this specification is that the real exchange rate is non-stationary if it was quite close to its long run equilibrium value in the last period and that there are driving forces like arbitrage that leads to mean-reversion if the real exchange rate departs from its long run equilibrium. Moreover, arbitrage may not be profitable if the departure is small. Therefore, the degree of mean-reversion is small as well and vice versa. These facts make this ESTAR specification quite attractive for modeling economic time series like real exchange and interest rates, unemployment rates and log dividend yields.

There are a lot of economic theories like Purchasing Power Parity (PPP), to name a highly debated one, that imply certain time series properties, i.e. the stationarity of real exchange rates. Often the unit root hypothesis, which contradicts PPP, is tested against stationarity with extant linear unit root tests. These tests have reduced power when the true data generating process exhibits non-linearities. Therefore, many recent empirical studies make use of non-linear unit root tests.

Regarding the ESTAR specification from above, a popular Dickey-Fuller-type test has been proposed by Kapetanios et al. (2003). However, this test assumes that the location parameter c in the smooth transition function is equal to zero. On the contrary, a lot of empirical studies on real exchange rates report significant estimates of c, cf. Michael et al. (1997), Sarantis (1999), Taylor et al. (2001) and more recently, Rapach and Wohar (2006). When relaxing this assumption, we are faced with a non-standard testing problem, i.e. a joint hypothesis where one parameter is one-sided under the alternative while all others are two-sided. Since standard inference techniques are not appropriate in this situation, we make use of the new approach by Abadir and Distaso (2007) who propose a class of modified test statistics in order to tackle such non-standard testing problems. Our aim is to derive a unit root test allowing for a non-zero location parameter c that can compete with the extant one of Kapetanios et al. (2003) in terms of power.

After introducing the ESTAR specification in more detail and presenting the existing test by Kapetanios et al. (2003) in section 4.2, the inference techniques by Abadir and Distaso (2007) and the new test are discussed in section 4.3. The non-standard limiting distribution of the test statistic is derived and consistency of the test is proven. Moreover, we show that the limiting distribution remains unchanged if we account for potential serial correlation in the error terms by augmenting the test regression with lags of the dependent variable. By means of a Monte Carlo study in section 4.4 we compare the small sample properties of both tests under a variety of conditions. The new test is correctly sized and quite often superior in terms of power. Both tests have lower but substantial power when the true data generating process is a logistic STAR model. However, the new test has generally higher power against logistic STAR models than the extant test. Finally, we provide an empirical application to a monthly real effective exchange rate time series for the European Union in section 4.5. The results suggest the validity of PPP if the new test is used and the opposite if the extant test is applied. Conclusions are drawn in section 4.6. Proofs are given in the Appendix of this chapter.

4.2 DF-type Unit Root Test against ESTAR

The ESTAR specification we are concerned with is formally given by

$$\Delta y_t = \alpha y_{t-1} + \phi y_{t-1}(1 - \exp\{-\gamma(y_{t-1} - c)^2\}) + \varepsilon_t \ ,$$

where $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$. If the smoothness parameter γ approaches zero, the ESTAR model becomes a linear AR(1) model, i.e. $\Delta y_t = \alpha y_{t-1} + \varepsilon_t$ that is stationary if $-2 < \alpha < 0$. In the following, α is set equal to zero which means that the ESTAR model becomes a random walk if $\gamma = 0$. Kapetanios et al. (2003) show that the ESTAR model under the restriction $\alpha = 0$,

$$\Delta y_t = \phi y_{t-1}(1 - \exp\{-\gamma(y_{t-1} - c)^2\}) + \varepsilon_t \ ,$$

is globally stationary if $-2 < \phi < 0$ is true although it is locally non-stationary in the sense that it contains a partial unit root when $y_{t-1} = c$ holds. Additionally note that the random walk model can also be achieved when imposing the restriction $\phi = 0$. This means that a direct test for the unit root hypothesis is infeasible since ϕ is not identified when testing $H_0 : \gamma = 0$ and vice versa.

A popular approach to avoid the presence of nuisance parameters under the null hypothesis is to use a Taylor approximation of the smooth transition function $G(y_{t-1}; \gamma, c) =$

$1 - \exp\{-\gamma(y_{t-1} - c)^2\})$ around $\gamma = 0$, see Luukkonen et al. (1988). This approach was adopted by Kapetanios et al. (2003) and we construct the new test on the same basis. More specifically, Kapetanios et al. (2003) make the restriction $c = 0$ and consider the model

$$\Delta y_t = \phi y_{t-1}(1 - \exp\{-\gamma y_{t-1}^2\}) + \varepsilon_t \ .$$

An application of a first-order Taylor approximation leads to the auxiliary regression

$$\Delta y_t = \beta_1 y_{t-1}^3 + u_t \ , \tag{4.1}$$

with u_t being a noise term depending on ε_t, ϕ and the remainder of the Taylor expansion. Obviously, it looks very much like the famous Dickey-Fuller test regression without deterministic terms. The cubic term y_{t-1}^3 approximates the ESTAR non-linearity.

The authors suggest a t-test for the unit root hypothesis against globally stationary ESTAR which corresponds to $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 < 0$. Hence, the unit root test is carried out by estimating the auxiliary regression (1) and computing a Dickey-Fuller-type t-test, labeled as KSS,

$$\text{KSS} \equiv \frac{\hat{\beta}_1}{\sqrt{\hat{\text{var}}(\hat{\beta}_1)}} = \frac{\sum_{t=1}^T y_{t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^T y_{t-1}^6}}, \tag{4.2}$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T (\Delta y_t - \hat{\beta}_1 y_{t-1}^3)^2$ is the usual estimator of the error variance. Let $W(r)$ be the Brownian motion defined on $r \in [0, 1]$ and let $\Rightarrow$ denote convergence in distribution. The limiting distribution of the KSS statistic is given by

$$\text{KSS} \Rightarrow \frac{\frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2 dr}{\left(\int_0^1 W(r)^6\right)^{1/2}} \ ,$$

see Theorem 1 in Kapetanios et al. (2003). Regarding deterministic terms, they suggest to de-mean or de-trend the data in a first step, i.e.

$$y_t = \omega' d_t + v_t$$

with $d_t = 1$ or $d_t = [1 \quad t]'$ and ω is a parameter vector of suitable dimension. In a second step, the unit root test is applied to $\widehat{v}_t$. As a consequence, the asymptotic distribution of the KSS statistic depends on functionals of a de-meaned or de-trended Brownian motion, respectively. The de-meaned and de-trended Brownian motion are given by

$$\begin{aligned} W(r) &- \int_0^1 W(r)dr \ , \\ W(r) &+ (6r-4)\int_0^1 W(r)dr + (12r-6)\int_0^1 rW(r)dr \ , \end{aligned}$$

respectively. For details concerning this test such as proofs and critical values see Kapetanios et al. (2003).

4.3 Modified Wald-type Test

In order to allow for a non-zero location parameter c in the exponential transition function we consider the non-linear time series model

$$\Delta y_t = \phi y_{t-1}(1 - \exp\{-\gamma(y_{t-1} - c)^2\}) + \varepsilon_t \ . \tag{4.3}$$

Following Kapetanios et al. (2003), we apply a first-order Taylor approximation to $G(y_{t-1}; \gamma, c) = (1 - \exp\{-\gamma(y_{t-1} - c)^2\})$ around $\gamma = 0$ and proceed with the test regression

$$\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1} + u_t \ .$$

Following Kapetanios et al. (2003) we impose $\beta_3 = 0$ to improve the power of the test, see Kapetanios et al. (2003), footnote 5. Thus, we proceed with

$$\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2 y_{t-1}^2 + u_t \ . \tag{4.4}$$

where $\beta_1 = \gamma\phi$ and $\beta_2 = -2c\gamma\phi$. We are interested in the pair of hypotheses given by $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$. In the test regression (4.4), this pair of hypothesis is equivalent to $H_0 : \beta_1 = \beta_2 = 0$ against $H_1 : \beta_1 < 0, \beta_2 \neq 0$. Note that the two-sidedness of β_2 under H_1 stems from the fact that c is allowed to take real values. This testing problem is non-standard in the sense that one parameter is one-sided under H_1 while the other one is two-sided. A standard Wald test would be inappropriate and we therefore apply the methods of Abadir and Distaso (2007) to derive a suitable test. In a nutshell, the one-sided parameter is orthogonalized with respect to the two-sided one. The modified Wald test builds upon the one-sided parameter (β_1) and the transformed two-sided parameter, say $\beta_2^{\perp}$, that are stochastically independent by definition.

Let the parameter vector of the regression model (4.4) be $\theta = [\beta_1 \ \ \beta_2]'$. Following the notation of Abadir and Distaso (2007), the null hypothesis of a unit root is rewritten as

$$H_0 : h(\theta) \equiv [h_1(\theta) \ \ h_2(\theta)]' = [\beta_1 \ \ \beta_2]' = [0 \ \ 0]' \ .$$

The alternative hypothesis of a globally stationary ESTAR model is given by

$$H_1 : h_1(\theta) < 0 \text{ or } h_2(\theta) \neq 0 \ ,$$

which includes the subset hypothesis $H_1^{\cap} : h_1(\theta) < 0$ and $h_2(\theta) \neq 0$. Theorem 6 in Abadir and Distaso (2007) states that the modified Wald test is consistent against H_1 as well as $H_1^{\cap}$. The standard Wald test statistic based on the Hessian matrix $\mathcal{H}$ is

$$W_{\mathcal{H}} = h(\hat{\theta})' V^{-1} h(\hat{\theta})$$

where $V \equiv \left[\frac{\partial h(\theta)}{\partial \theta'}\Big|_{\theta=\hat{\theta}} (-\mathcal{H})^{-1} \frac{\partial h(\theta)'}{\partial \theta}\Big|_{\theta=\hat{\theta}} \right]$ with elements v_{ij}. In general, the modified Wald test statistic of Abadir and Distaso (2007) is given by

$$\begin{aligned} \tau &= \left(\frac{\partial h_{2\cdot 1}(\theta)}{\partial \theta'}\Big|_{\theta=\hat{\theta}} (-\mathcal{H})^{-1} \frac{\partial h_{2\cdot 1}(\theta)'}{\partial \theta}\Big|_{\theta=\hat{\theta}} \right)^{-1} \hat{h}_{2\cdot 1}(\hat{\theta})^2 \\ &+ \ 1(h_1(\hat{\theta}) < 0) \left(\frac{\partial h_1(\theta)}{\partial \theta'}\Big|_{\theta=\hat{\theta}} (-\mathcal{H})^{-1} \frac{\partial h_1(\theta)'}{\partial \theta}\Big|_{\theta=\hat{\theta}} \right)^{-1} \hat{h}_1(\hat{\theta})^2 \ , \end{aligned}$$

with $h_{2\cdot 1}(\theta)$ being

$$h_{2\cdot 1}(\theta) = h_2(\theta) - \frac{h_1(\theta) v_{21}}{v_{11}} \ .$$

The estimator of $h_{2\cdot 1}(\theta)$ is simply given by $\hat{h}_{2\cdot 1}(\hat{\theta}) = h_2(\hat{\theta}) - \frac{h_1(\hat{\theta})\hat{v}_{21}}{\hat{v}_{11}}$. Based on these results, straightforward calculations lead us to

$$\tau = \left(\hat{v}_{22} - \frac{\hat{v}_{21}^2}{\hat{v}_{11}} \right)^{-1} \left(\hat{\beta}_2 - \hat{\beta}_1 \frac{\hat{v}_{21}}{\hat{v}_{11}} \right)^2 + 1(\hat{\beta}_1 < 0) \frac{\hat{\beta}_1^2}{\hat{v}_{11}} \ ,$$

which is the new test statistic for the unit root hypothesis against globally stationary ESTAR. A simpler and more intuitive way to formulate this statistic is

$$\tau = t^2_{\beta_2^{\perp}=0} + 1(\hat{\beta}_1 < 0) t^2_{\beta_1=0} \ .$$

The two summands appearing in the test statistic τ can be interpreted as follows: the first term is a squared t-statistic for the hypothesis $\beta_2^{\perp} \equiv \beta_2 - \beta_1 v_{21}/v_{11} = 0$ with $\beta_2^{\perp}$ being orthogonal to β_1. Additionally, the second term is a squared t-statistic for the hypothesis $\beta_1 = 0$, the one-sidedness under H_1 is achieved by the multiplied indicator function. In the next step, the limiting distribution of τ is derived.

Assumption 1 *y_t is a random walk, i.e. $y_t = y_{t-1} + \varepsilon_t$ with $y_0 = 0$ and u_t being i.i.d.$(0, \sigma^2)$ and $E|u_t|^\delta < \infty$ for $\delta \geq 6$.*

In Theorem 5 we derive the asymptotic distribution of τ under the null hypothesis $H_0 : \gamma = 0$.

Theorem 5. *Under Assumption 1 the τ statistic has the following asymptotic distribution which is free of nuisance parameters:*

$$\tau \Rightarrow \mathcal{A}(W(r)) + \mathcal{B}(W(r)) \; ,$$

where $\mathcal{A}$ and $\mathcal{B}$ are functions of the Brownian motion $W(r)$ that are given in the proof. Under the alternative hypothesis $H_1 : \gamma > 0$ the τ statistic diverges with rate T.

Proof 1. *See Appendix of this chapter.*

We follow the approach of Kapetanios et al. (2003) and de-mean or de-trend the data in a first step when allowing for deterministic terms. This means that the Brownian motion $W(r)$ appearing in the limiting distribution of the τ statistic has to be replaced by a de-meaned or de-trended Brownian motion, respectively.

Next, we consider the case of serially correlated errors. We allow for stationary linear innovations that are generated by a short-memory process v_t.

Assumption 2 *$v_t = \psi(L)u_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$, where $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ and $u_t \sim$ i.i.d.$(0, \sigma^2)$.*

Table 4.1: Critical Values

	$d_t = 0$	$d_t = 1$	$d_t = [1 \quad t]'$
1%	13.15	13.75	17.10
5%	9.53	10.17	12.82
10%	7.85	8.60	11.10

In Theorem 6 we show that the asymptotic distribution of the τ statistic does not change when adding a sum of lagged differences on the right hand side of the test regression, i.e.

$$\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2 y_{t-1}^2 + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + u_t \ . \tag{4.5}$$

Theorem 6. *The asymptotic distribution of the τ statistic does not change when the test regression in (4.5) is used instead of (4.4).*

Proof 2. *See Appendix of this chapter.*

Alternatively, one could consider more general error processes and derive a Phillips-Perron-type test for the unit root hypothesis, see Rothe and Sibbertsen (2006) and Sandberg (2008). However, we focus on the augmented Dickey-Fuller version in this chapter and leave the other for future research.

4.4 Monte Carlo Study

This section covers the Monte Carlo study that compares the small sample performance of the new unit root test and the existing test by Kapetanios et al. (2003). Throughout this section we set the number of observations T equal to 300 which is a reasonable sample size for many macroeconomic and financial time series like unemployment rates and interest rates. Furthermore, 500 initial observations are deleted to reduce the effect of initial conditions.

Asymptotic critical values for the modified Wald-type test τ are provided in Table 4.1. They are based on 20,000 replications and $T = 1,000$. We report critical values for raw

($d_t = 0$), de-meaned ($d_t = 1$) and de-trended data ($d_t = [1 \quad t]'$) for nominal significance levels of one, five and ten percent, respectively.

We investigate the size of both tests under the following data generating processes

$$y_t = y_{t-1} + \varepsilon_t \text{ with } \varepsilon_t = u_t \tag{4.6}$$

$$y_t = y_{t-1} + \varepsilon_t \text{ with } \varepsilon_t = \rho\varepsilon_{t-1} + u_t \tag{4.7}$$

$$y_t = y_{t-1} + \varepsilon_t \text{ with } \varepsilon_t = u_t - \theta u_{t-1} , \tag{4.8}$$

where u_t is drawn from the standard normal distribution. The errors ε_t follow an i.i.d. process, an AR(1) or MA(1) process. We adopt the approach of Phillips and Sul (2003) and sample parameters from uniform distributions in order to cover a wide range of values in a relatively small number of experiments. Hence, the autoregressive and moving average parameters ρ and θ are drawn from uniform distributions:

$$\rho \sim U\left[\underline{\rho}, \bar{\rho}\right] \text{ and } \theta \sim U\left[\underline{\theta}, \bar{\theta}\right] .$$

We specify $\underline{\rho} = \underline{\theta} = 0$ and $\bar{\rho} = \bar{\theta} = 0.4$. The size experiments are based on 5,000 replications, results can be found in Table 4.2. Most rejection rates under the null hypothesis are quite close to the nominal ones which suggests that both tests are correctly sized. We observe that both tests are a little bit oversized in the presence of errors that follow a first-order moving average process.

Next, we study the power of both tests by considering different settings in the globally stationary non-linear ESTAR process for the parameters c and γ. The data generating process we consider under the non-linear ESTAR alternative is

$$\Delta y_t = \phi y_{t-1}(1 - \exp\{-\gamma(y_{t-1} - c)^2\}) + \varepsilon_t \tag{4.9}$$

with $\phi = -1$. This restriction is often imposed in empirical studies, see for example Taylor et al. (2001). The location parameter c is set either equal to zero or it is drawn from a uniform distribution with lower and upper bound, $\underline{c}$ and $\bar{c}$, respectively. Analogously,

Table 4.2: Size Experiments

d_t	Test	i.i.d.			AR(1)			MA(1)		
		1.0	5.0	10.0	1.0	5.0	10.0	1.0	5.0	10.0
0	KSS	0.9	5.1	9.9	1.0	5.0	10.0	1.2	4.7	10.1
	τ	1.1	5.2	10.9	1.0	5.0	10.0	1.7	5.2	11.9
1	KSS	1.1	5.1	9.9	1.1	4.8	9.7	1.3	6.0	11.8
	τ	1.1	4.9	9.8	0.9	4.6	9.5	1.2	5.6	11.8
$[1 \quad t]'$	KSS	1.1	4.9	10.2	0.8	5.2	9.6	1.4	6.5	10.9
	τ	1.0	5.2	10.1	0.9	5.4	10.3	1.5	6.6	11.2

Notes: Reported values are rejection rates of KSS and τ test under the validity of H_0.

the smoothness parameter γ is drawn from a uniform distribution with lower and upper bound $\underline{\gamma}$ and $\overline{\gamma}$, respectively:

$$c \sim U\left[\underline{c}, \bar{c}\right] \text{ and } \gamma \sim U\left[\underline{\gamma}, \bar{\gamma}\right] .$$

Results of these power experiments are reported in the upper panel of Table 4.3. In the first experiment we specify a zero location parameter ($c_0 \equiv c = 0$) and slow transition between regimes ($\gamma_l \equiv \gamma \sim U[0.001, 0.01]$). In the second and third experiments a non-zero location parameter is allowed by drawing it from a uniform distribution with lower and upper bound of -5 (-10) and 5 (10), respectively. The fourth and fifth settings restrict the upper bound $\bar{c}$ to zero in order to have a non-zero mean of c. Please note that we do not report results for experiments where the lower bound $\underline{c}$ is restricted to zero because there is no qualitative difference due to symmetry. The last two experiments are like the two previous ones but with fast transition between regimes, i.e. $\gamma_h \equiv \gamma \sim U[0.01, 0.1]$.

When interpreting the simulated rejection probabilities against ESTAR we observe that the new test is generally superior to the Kapetanios et al. (2003) test in terms of power. Only in some cases where the unit root tests are applied to raw data ($d_t = 0$), the KSS test performs somewhat better than the modified Wald test. Most applications of unit root tests in economics involve deterministic terms. When data is de-meaned or

Table 4.3: Power Experiments

d_t	Test	c_0, γ_l	$c_{\pm 5}, \gamma_l$	$c_{\pm 10}, \gamma_l$	c_{-5}, γ_l	c_{-10}, γ_l	c_{-5}, γ_h	c_{-10}, γ_h
				Exponential STAR				
0	KSS	99.6	92.4	73.0	92.0	74.4	97.0	97.2
	τ	95.3	88.1	72.6	88.5	74.5	98.2	97.9
1	KSS	91.1	87.6	77.8	88.4	79.7	98.1	97.5
	τ	92.9	92.3	93.3	91.8	93.9	100	100
$[1 \quad t]'$	KSS	77.5	74.2	64.1	73.4	65.9	97.7	96.3
	τ	81.6	78.9	78.4	79.6	78.7	100	100
				Logistic STAR				
0	KSS	97.8	95.4	87.8	94.9	89.2	99.9	99.9
	τ	83.2	82.5	82.7	80.3	83.3	99.9	100
1	KSS	75.6	72.4	63.7	71.4	64.6	99.0	94.4
	τ	79.0	77.5	78.3	77.9	80.0	100	100
$[1 \quad t]'$	KSS	53.3	50.4	46.5	50.8	48.5	97.5	92.3
	τ	58.6	54.9	57.7	55.4	59.4	99.5	99.8

Notes: Reported values are rejection rates of KSS test (upper entries) and τ test (lower entries). Nominal significance level is five percent.

de-trended, power gains up to 15 percent can be achieved by applying the new test.

In addition, we study the power of both tests against globally stationary logistic STAR (LSTAR) models. As noted by Kapetanios et al. (2003) a non-linear adjustment scheme alternative to the exponential one is a logistic smooth transition function. We use the second-order logistic function

$$G(y_{t-1}; \gamma, c_1, c_2) = 2/[1 + \exp(-\gamma(y_{t-1} - c_1)(y_{t-1} - c_2))] - 1$$

that has two location parameters, namely c_1 and c_2. Like the exponential smooth transition function it becomes constant if $\gamma \to 0$, which means that the non-linear logistic smooth transition model becomes linear. Without loss of generality, we set $c_1 = 0$ and draw c_2 from the uniform distribution as done before in the case of an exponential smooth transition. The data generating process is now given by

$$\Delta y_t = y_{t-1}(1 - 2/[1 + \exp(-\gamma(y_{t-1} - c_1)(y_{t-1} - c_2))]) + \varepsilon_t \ . \tag{4.10}$$

Empirical rejection frequencies are reported in the lower panel of Table 4.3. Both tests have higher power against ESTAR than against LSTAR models which is not surprising since both have the former one as specific alternative. Nonetheless, one might expect that both tests have substantial power against logistic STAR models because the Taylor approximation of a logistic STAR model is quite similar. Thus, a rejection of the null hypothesis does not necessarily contain information about the specific *type* of non-linear adjustment. When comparing both unit root tests, we come to the same conclusions as before. In addition we observe that the power is lower for de-trended data than for de-meaned data which is due to an additional parameter that has to be estimated when de-trending the data.

In sum, the new test shows good overall performance and is quite often more powerful than the existing test by Kapetanios et al. (2003), especially when the test is applied to de-meaned or de-trended data, which are the most important cases in practice.

4.5 Empirical Application

Unit root tests have become a very popular tool in the literature that is concerned with testing the validity of the Purchasing Power Parity (PPP) which is one of the most important parities in international macroeconomics. One can say that PPP holds if and only if the real exchange rate is stationary. Thus, testing the unit root hypothesis means testing the non-validity of the PPP theory. Since linear unit root tests like the ones of Dickey-Fuller (1979) and Phillips and Perron (1988) often fail to reject the null hypothesis of non-stationarity when being applied to real exchange rate data, researchers tend to use non-linear unit root tests where the specific model that is true under the alternative is congruent with economic models of financial markets. For example, STAR models for the real exchange rate can be interpreted in the context of transaction costs and arbitrage, see Dumas (1992), Sercu et al. (1995) and Michael et al. (1997).

However, rejecting the null hypothesis in favor of a non-linear alternative while a linear

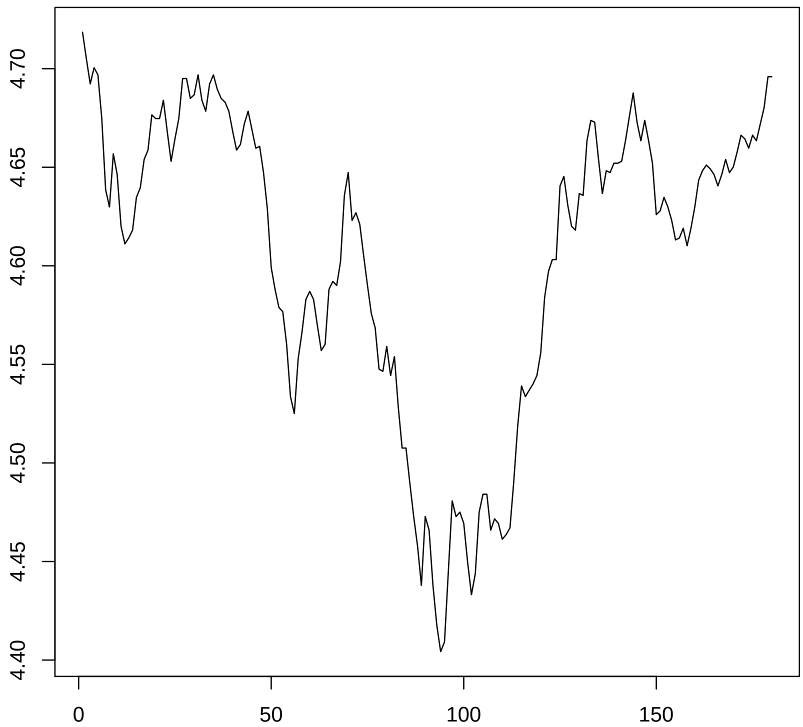


Figure 4.1: Logarithm of Real Effective Exchange Rate (January, 1993 to December, 2007).

Dickey-Fuller test does not reject in favor of a linear alternative might hint at non-linearities. Nonetheless, one should be careful with the conclusion that non-linearity is of ESTAR-type because the test regression approximates a lot of non-linear models and such tests can have substantial power against other non-linear mean-reverting processes, see section 4.4.

We apply both unit root tests against non-linear alternatives and two famous tests against linear alternatives to the monthly real effective exchange rate time series for the European Union. Our data is taken from Datastream (code: EMXTW..RF) and spans from 1993:01 to 2007:12 implying 180 observations. The logged time series is depicted in Figure 4.1. No linear trend can be seen in the data but the mean appears to be highly significant. Hence, we de-mean the data in a first step. In a second step we estimate the test regressions with a lag length chosen accordingly to the Schwarz information criterion ($\widehat{p} = 1$). We obtain KSS $= -2.21$ which is not significant at the ten percent level suggesting that PPP does not hold. On the contrary, using the new test one has to reject the null hypothesis since $\tau = 9.19$ is significant at the ten percent level which indicates that PPP holds true. Furthermore, the unit root tests against linear alternatives by Dickey and Fuller (1979) (DF) and Phillips and Perron (1988) (PP) do not provide any evidence against the null hypothesis. The test statistics are DF $= -1.76$ and PP $= -1.60$, respectively.

We conclude that non-linearities, potentially of exponential STAR-type with non-zero location, are present in the data but that they are not detected by applying existing tests. The modified Wald unit root test yields new evidence on the stationarity of the EU real effective exchange rate which suggests the validity of PPP.

4.6 Conclusions

This chapter contributes to the literature on non-linear unit root tests by generalizing the existing test by Kapetanios et al. (2003) with respect to a non-zero location parameter. The resulting non-standard testing problem is tackled by deriving a modified Wald test that builds upon the inference techniques by Abadir and Distaso (2007). The non-standard limiting distribution of the test statistic has been derived under standard assumptions. The Monte Carlo study shows that the new test is superior to the extant test in most situations. An empirical application to the EU real effective exchange rate underpins its usefulness.

4.7 Appendix

Proof of Theorem 5: In order to simplify the notation, we write $\int$ instead of $\int_0^1$ in the following. The proof makes use of the following convergence results, see Hansen (1992) and Hamilton (1994). We have

$$\begin{aligned}
\frac{1}{T^{(i+2)/2}}\sum_{t=1}^{T} y_{t-1}^i &\Rightarrow \sigma^i \int W(r)^i dr \qquad \text{for } i=4,5,6\\
\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}^2\Delta y_t &\Rightarrow \sigma^3\left(\frac{1}{3}W(1)^3-\int W(r)dr\right)\\
\frac{1}{T^{2}}\sum_{t=1}^{T} y_{t-1}^3\Delta y_t &\Rightarrow \sigma^4\left(\frac{1}{4}W(1)^4-\frac{3}{2}\int W(r)^2dr\right) .
\end{aligned}$$

We first note that the second summand of τ is given by $1(\hat\beta_1<0)t^2_{\beta_1=0}$ and that the OLS estimator for β_1 in (4.4) is given by

$$\hat\beta_1=\frac{\sum_{t=1}^T y_{t-1}^4\sum_{t=1}^T y_{t-1}^3\Delta y_t-\sum_{t=1}^T y_{t-1}^2\Delta y_t\sum_{t=1}^T y_{t-1}^5}{\sum_{t=1}^T y_{t-1}^4\sum_{t=1}^T y_{t-1}^6-\left(\sum_{t=1}^T y_{t-1}^5\right)^2} .$$

Under Assumption 1 and by using the convergence results from above we obtain $\hat\beta_1 \xrightarrow{P} \beta_1$ and $\hat\beta_1=O_P(T^{-2})$. Furthermore,

$$t_{\beta_1=0}=\frac{\sum_{t=1}^T y_{t-1}^4\sum_{t=1}^T y_{t-1}^3\Delta y_t-\sum_{t=1}^T y_{t-1}^2\Delta y_t\sum_{t=1}^T y_{t-1}^5}{\sqrt{\hat\sigma^2\left(\left(\sum_{t=1}^T y_{t-1}^4\right)^2\sum_{t=1}^T y_{t-1}^6-\sum_{t=1}^T y_{t-1}^4\left(\sum_{t=1}^T y_{t-1}^5\right)^2\right)}} .$$

Again, by using the convergence results it follows that

$$t_{\beta_1=0}\Rightarrow\frac{\left(\int W(r)^4dr\right)\left(\frac14 W(1)^4-\frac32\int W(r)^2dr\right)-\left(\frac13 W(1)^3-\int W(r)dr\right)\left(\int W(r)^5dr\right)}{\sqrt{\left(\int W(r)^4dr\right)^2\left(\int W(r)^6dr\right)-\left(\int W(r)^4dr\right)\left(\int W(r)^5dr\right)^2}} ,$$

and by applying the CMT it follows directly that $t^2_{\beta_1=0}$ converges to the square of the previous function which gives an expression for $\mathcal{B}(W(r))$. Regarding the first summand of τ, we have for the nominator of $t_{\beta_2^\perp=0}$

$$\hat\beta_2-\hat\beta_1\frac{\hat v_{21}}{\hat v_{11}}=\frac{\sum_{t=1}^T y_{t-1}^4\sum_{t=1}^T y_{t-1}^6\sum_{t=1}^T y_{t-1}^2\Delta y_t-\sum_{t=1}^T y_{t-1}^2\Delta y_t\left(\sum_{t=1}^T y_{t-1}^5\right)^2}{\left(\sum_{t=1}^T y_{t-1}^4\right)^2\sum_{t=1}^T y_{t-1}^6-\sum_{t=1}^T y_{t-1}^4\left(\sum_{t=1}^T y_{t-1}^5\right)^2} .$$

For the denominator of $t_{\beta_2^\perp=0}$ we have after simple calculations

$$\sqrt{\hat{\upsilon}_{22}-\frac{\hat{\upsilon}_{21}^2}{\hat{\upsilon}_{11}}}=\sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^T y_{t-1}^4}}\ .$$

Therefore

$$t_{\beta_2^\perp=0}=\frac{\sum_{t=1}^T y_{t-1}^2\Delta y_t\left(\left(\sum_{t=1}^T y_{t-1}^4\right)^{3/2}\sum_{t=1}^T y_{t-1}^6-\left(\sum_{t=1}^T y_{t-1}^4\right)^{1/2}\left(\sum_{t=1}^T y_{t-1}^5\right)^2\right)}{\hat{\sigma}\left(\left(\sum_{t=1}^T y_{t-1}^4\right)^2\sum_{t=1}^T y_{t-1}^6-\sum_{t=1}^T y_{t-1}^4\left(\sum_{t=1}^T y_{t-1}^5\right)^2\right)}$$

Using the convergence results stated above it follows that

$$t_{\beta_2^\perp=0}\Rightarrow\frac{\left(\frac{1}{3}W(1)^3-\int W(r)dr\right)\left(\left(\int W(r)^4dr\right)^{3/2}\left(\int W(r)^6dr\right)-\left(\int W(r)^4dr\right)^{1/2}\left(\int W(r)^5dr\right)^2\right)}{\left(\int W(r)^4dr\right)^2\left(\int W(r)^6dr\right)-\left(\int W(r)^4dr\right)\left(\int W(r)^5dr\right)^2}$$

Again, by CMT it follows that $t^2_{\beta_2^\perp=0}$ converges in distribution to the square of the previous function which gives an expression for $\mathcal{A}(W(r))$. It is easy to show that $\hat{\sigma}^2\xrightarrow{p}\sigma^2$, see Kapetanios et al. (2003). Under the alternative hypothesis, Δy_t, y_{t-1}^2 and y_{t-1}^3 are $I(0)$ processes and it is easy to show that the terms appearing in the test statistic are $O_P(T)$. Then, $t^2_{\beta_2^\perp=0}=O_P(T)$ and $t^2_{\beta_1=0}=O_P(T)$, therefore $\tau=O_P(T)$. The τ statistic is therefore diverging with rate T. $\square$

Proof of Theorem 6: The proof is very similar to the one of Kapetanios et al. (2003) as it uses the same arguments. Let

$$Z=[\Delta y_{-1},\Delta y_{-2},\ldots,\Delta y_{-p}]$$

with

$$\Delta y_{-i}=[\Delta y_{-i+1},\Delta y_{-i+2},\ldots,\Delta y_{T-i}]$$

and $M_T=I_T-Z(Z'Z)^{-1}Z$. Note that, $\hat{\sigma}^2=\frac{1}{T}\varepsilon' M_T\varepsilon=\frac{1}{T}\varepsilon'\varepsilon+o_p(1)\xrightarrow{p}\sigma^2$ with $\varepsilon=[\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_T]'$. Moreover, we have

$$\frac{1}{T^{(i+2)/2}}y_{-1}^{i/2'}M_T y_{-1}^{i/2}=\frac{1}{T^{(i+2)/2}}y_{-1}^{i/2'}y_{-1}^{i/2}+o_p(1)\ \Rightarrow\ \lambda^i\int W(r)^i dr\quad\text{for } i=4,6$$
$$\frac{1}{T^{7/2}}y_{-1}^{5/2'}M_T y_{-1}^{5/2}=\frac{1}{T^{7/2}}y_{-1}^{5/2'}y_{-1}^{5/2}+o_p(1)\ \Rightarrow\ \lambda^5\int W(r)^5 dr\ ,$$

and additionally,

$$\begin{aligned} \frac{1}{T^{3/2}} y_{-1}^{2'} M_T \varepsilon = \frac{1}{T^{3/2}} y_{-1}^{2'} \varepsilon + o_p(1) \quad &\Rightarrow \quad \frac{1}{3} \lambda^3 W(1)^3 - \lambda\sigma^2 \int W(r)dr \\ \frac{1}{T^2} y_{-1}^{3'} M_T \varepsilon = \frac{1}{T^2} y_{-1}^{3'} \varepsilon + o_p(1) \quad &\Rightarrow \quad \frac{1}{4} \lambda^4 W(1)^4 - \frac{3}{2} \sigma^2 \lambda^2 \int W(r)^2 dr \ , \end{aligned}$$

where λ^2 is the long-run variance of Δy_t under the null hypothesis. Based on these results we have

$$\begin{aligned} t_{\beta_1=0} &= \frac{\left(y_{-1}^{2'} M_T y_{-1}^2\right)\left(y_{-1}^{3'} M_T \varepsilon\right) - \left(y_{-1}^{2'} M_T \varepsilon\right)\left(y_{-1}^{5/2'} M_T y_{-1}^{5/2}\right)}{\sqrt{\hat{\sigma}^2 \left(\left(y_{-1}^{2'} M_T y_{-1}^2\right)^2 \left(y_{-1}^{3'} M_T y_{-1}^3\right) - \left(y_{-1}^{2'} M_T y_{-1}^2\right)\left(y_{-1}^{5/2'} M_T y_{-1}^{5/2}\right)^2\right)}} \\ &= \frac{\left(y_{-1}^{2'} y_{-1}^2\right)\left(y_{-1}^{3'} \varepsilon\right) - \left(y_{-1}^{2'} \varepsilon\right)\left(y_{-1}^{5/2'} y_{-1}^{5/2}\right)}{\sqrt{\hat{\sigma}^2 \left(\left(y_{-1}^{2'} y_{-1}^2\right)^2 \left(y_{-1}^{3'} y_{-1}^3\right) - \left(y_{-1}^{2'} y_{-1}^2\right)\left(y_{-1}^{5/2'} y_{-1}^{5/2}\right)^2\right)}} + o_p(1) \ . \end{aligned}$$

Furthermore,

$$\begin{aligned} t_{\beta_2^{\perp}=0} &= \frac{y_{-1}^{2'} M_T \varepsilon \left(\left(y_{-1}^{2'} M_T y_{-1}^2\right)^{3/2} \left(y_{-1}^{3'} M_T y_{-1}^3\right) - \left(y_{-1}^{2'} M_T y_{-1}^2\right)^{1/2} \left(y_{-1}^{5/2'} M_T y_{-1}^{5/2}\right)^2\right)}{\hat{\sigma} \left(\left(y_{-1}^{2'} M_T y_{-1}^2\right)^2 \left(y_{-1}^{3'} M_T y_{-1}^3\right) - \left(y_{-1}^{2'} M_T y_{-1}^2\right)\left(y_{-1}^{5/2'} M_T y_{-1}^{5/2}\right)^2\right)} \\ &= \frac{y_{-1}^{2'} \varepsilon \left(\left(y_{-1}^{2'} y_{-1}^2\right)^{3/2} \left(y_{-1}^{3'} y_{-1}^3\right) - \left(y_{-1}^{2'} y_{-1}^2\right)^{1/2} \left(y_{-1}^{5/2'} y_{-1}^{5/2}\right)^2\right)}{\hat{\sigma} \left(\left(y_{-1}^{2'} y_{-1}^2\right)^2 \left(y_{-1}^{3'} y_{-1}^3\right) - \left(y_{-1}^{2'} y_{-1}^2\right)\left(y_{-1}^{5/2'} y_{-1}^{5/2}\right)^2\right)} + o_p(1) \ , \end{aligned}$$

which, as we have shown before, has the asymptotic distribution given in Theorem 5. Finally, among similar lines in the foregoing proof, it is easily seen that the τ test is consistent under the alternative. □

Chapter 5

What do we know about Real Exchange Rate Non-linearity?

Co-authored with Michael Frömmel, Lukas Menkhoff and Philipp Sibbertsen

5.1 Introduction

The debate about real exchange rate behavior suggests that Purchasing Power Parity (PPP) may hold as a longer run concept (e.g. Taylor and Taylor, 2004). Our understanding of exchange rate dynamics that bring about longer run PPP, however, is much less clear. In particular, there are two competing approaches in wider use aiming for modeling a non-linear adjustment towards PPP, i.e. Exponential Smooth Transition Autoregressive (ESTAR) models and Markov Switching models (see e.g. Michael et al. 1997, and Kanas, 2006). This naturally raises the question which one of these approaches is more appropriate? Does real exchange rate dynamics rather follow an ESTAR or a Markov Switching autoregressive (MSAR) process which implies more or less tendency towards PPP, respectively? Or is there no adjustment at all which would be implied by the presence of a unit root?

We contribute to the empirical literature analyzing both models in a comparative perspective by using extant and introducing new thorough statistical unit root testing procedures. The empirical power of different non-linear unit root tests is examined under real world parameter constellations. These are obtained from Taylor et al. (2001) for ESTAR models and from own calculations for MSAR models which are both fitted to major monthly real exchange rates, namely to the German Mark (DEM), the Japanese Yen (JPY), the British Pound (GBP) against the US Dollar (USD). Fortunately, these time series also cover some variety in the parameters and persistence properties as we show later. By Monte Carlo simulations, we generate realizations of these data generating processes under the assumption that PPP holds and compute the power of the Kapetanios et al. (2003) and the Park and Shintani (2005) unit root tests against the ESTAR alternative - in short: ESTAR tests - as well as the power of a newly developed unit root test against a MSAR process. We report the power of these tests when the alternative is true (this is the true DGP is ESTAR/MSAR for the ESTAR/MSAR test) and when the alternative is misspecified (this is the true DGP is MSAR/ESTAR for the ESTAR/MSAR test).

We find that ESTAR tests have low empirical power so that evidence for ESTAR processes is weak at best, whereas the newly developed unit root test against Markov Switching is powerful. This implies that disequilibrium forces in real exchange rates - as captured by Markov Switching processes - may be stronger than seen before. In addition, a non-rejection of an ESTAR test does not mean that we have to reject the Purchasing Power Parity hypothesis completely. Furthermore, it turns out that the power of the ESTAR tests is even slightly higher under the misspecified alternative compared to the true alternative. Thus, rejecting the Null of a unit root by applying an ESTAR test does not allow the premature conclusion that the non-linearity of the true underlying data generating process is really of an ESTAR-type. It is more likely for the model to be any other non-linear process such as Markov Switching. The situation for our unit root test in a Markov Switching framework is much better: on the one hand, its power is substantially higher against the true alternative and on the other hand, it appears to be robust against ESTAR, meaning that it has quite a low power. This implies that the probability of confusing both processes is generally very low when applying the proposed Markov Switching test.

This research into the power of unit root tests has an obvious statistical motivation. In addition, it also has an intuitive economic motivation as ESTAR and Markov Switching processes imply a different understanding of the foreign exchange market. In short and somewhat overstating the point, the ESTAR view of real exchange rates emphasizes the tendency towards PPP and thus towards long run equilibrium, whereas the Markov Switching view emphasizes the fact that there may also be forces driving real exchange rates away from equilibrium and thus causing bubbles. Taylor (2005) and similarly De Grauwe and Grimaldi (2005) link the reasoning behind these views to heterogenous actors in this market, i.e. international goods arbitrage and short-term speculation.

ESTAR models, such as Taylor et al. (2001), pick up the argument that due to taxes and transportation costs goods are not traded internationally as long as the price levels do not differ too much between different countries (see the model in Dumas, 1992).[1] Therefore, the real exchange rate behaves like a random walk when it is close to its equilibrium value. As soon as the price differences increase, a smooth transition process starts and arbitrage will adjust prices and thus the real exchange rate towards PPP. This behavior can be well described by an ESTAR model with a unit root regime switching to an autoregressive regime if the process departs from its equilibrium.

In contrast, Markov Switching models, in particular the Markov Switching model as applied in Kanas (2006), argue that real exchange rates may be driven by various forces, some stabilizing - hinting at goods arbitrage - and some destabilizing - having short-term

[1] ESTAR models build on the STAR model of Teräsvirta (1994). Applications to foreign exchange include Michael et al. (1997) who test this model with interwar data for several exchange rates as well as with the long-span data on GBP/USD of Lothian and Taylor (1996), whereas Taylor et al. (2001) test the model at four exchanges rates against the USD during the post war period.

speculation in mind.[2] Thus, one regime may be more stabilizing and akin to the ESTAR view, whereas the other regime is different, i.e. either less stabilizing or even dominated by short term speculators which may cause exchange rate bubbles. The main difference is that the switch between the regimes is not necessarily linked to the degree of deviation from PPP.

Both views of real exchange rate behavior have a strong substantiation in international finance research. The ESTAR view is tentatively supported by long lasting research on PPP which has yielded the insight that forces towards PPP have been underestimated in earlier studies (survey in Sarno and Taylor, 2002). However, also the Markov Switching view has remarkable economic substantiation in models of heterogeneous agents in the foreign exchange market, such as early Frankel and Froot (1990) or recently De Grauwe and Grimaldi (2006). Although these views do not overwhelmingly imply a policy stance, the first view will clearly tend towards more benign neglect of the foreign exchange market than the latter view (further implications are discussed by Sarno, 2005, p.685f.).

Our application of unit root tests to real exchange rates shows, indeed, that the ESTAR tests cannot reject the unit root. This implies that either PPP does not hold or that the non-linear alternative to the unit root does not capture exchange rate properties well enough. It is consequently revealing that the Markov Switching test rejects the Null of a unit root in four out of the six major real exchange rates considered. This indicates that real exchange rate dynamics may be well characterized by Markov Switching processes.

The paper is organized as follows. Section 5.2 introduces the ESTAR model and the

[2]The Markov Switching model introduced by Hamilton (1989) was first applied to nominal exchange rates by Engel and Hamilton (1990) and in different settings by Engel (1994), Cheung and Erlandsson (2004) and Frömmel et al. (2005). The specific form of a Markov Switching error correction model we are interested in here, i.e. a Markov Switching error correction model applied to real exchange rates, has been introduced by Hall et al. (1997) and Psaradakis et al. (2004), first applied to our problem by Kanas (2006).

considered tests in more detail. In section 5.3 we propose a new unit root test against a MSAR process. Section 5.4 contains our Monte Carlo study, section 5.5 applies the unit root tests introduced above on six real exchange rates and section 5.6 concludes.

5.2 Unit Root Tests against ESTAR

In this section we briefly review the ESTAR model and the tests for a unit root against the ESTAR alternative which are applied in this paper. The ESTAR model we consider in our work, as used in several studies like Michael et al. (1997), Sarantis (1999), Taylor et al. (2001) and more recently, Rapach and Wohar (2006), is defined by

$$\Delta y_t = \phi y_{t-1} G(z_t; \gamma, c) + \varepsilon_t \tag{5.1}$$

where ε_t is assumed to be a zero mean white noise process and the autoregressive parameter is restricted to $\phi = -1$. The bounded exponential smooth transition function G depends on the transition variable z_t, the smoothness parameter $\gamma > 0$ and the location parameter c:

$$G(y_{t-1}; \gamma, c) = 1 - \exp\{-\gamma(z_t - c)^2\} \in [0, 1] \ .$$

In the following, we set $z_t = y_{t-1}$ which is a common choice in the related literature. The parameter γ controls the speed of a regime switch, a higher value for γ implies a higher speed of transition from one regime to another. The location parameter c is the root of the transition function G which implies that y_t is locally non-stationary, since $\Delta y_t = \varepsilon_t$ if $y_{t-1} = c$. As long as $\phi < 0$, y_t is globally stationary although it has a partial unit root, see Kapetanios et al. (2003). We consider a globally stationary ESTAR model with two regimes: a unit root regime ($y_{t-1} = c$) and a symmetric mean-reverting regime ($y_{t-1} \gtrless c$). Note, that the mean of y_t and the variance of the error term ε_t are usually assumed to be constant across regimes.

The non-linear ESTAR model becomes a linear random walk if $\gamma \to 0$. In the case that $\phi = 0$ or $\gamma \to \infty$ holds, the ESTAR model also becomes a random walk. Therefore, testing the unit root hypothesis against ESTAR is complicated due to unidentified parameters under the null hypothesis which is known as the Davies problem, see Davies (1987). Usually, the testing problem $H_0 : \gamma = 0$ vs. $H_1 : \gamma > 0$ is considered.

Only a few tests has been proposed for the inference problem of testing the Null that the data generating process is a linear process with a unit root against the alternative of a stationary ESTAR process. Choi and Moh (2007) show via an extensive Monte Carlo study that the power of the linear Dickey-Fuller test has less power than unit root tests that are designed for non-linear alternatives. The following subsections briefly review two unit root tests against ESTAR which are studied in section 5.4 by Monte Carlo methods.

5.2.1 Dickey-Fuller-type Test

Kapetanios et al. (2003) suggest a modification of the Dickey-Fuller test for testing $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$. They make the simplifying assumption that $c = 0$ and the ESTAR model is, therefore, given by

$$\Delta y_t = \phi y_{t-1}(1 - \exp\{-\gamma y_{t-1}^2\}) + \varepsilon_t \ .$$

Note that ϕ is unidentified under the null hypothesis $H_0 : \gamma = 0$. Luukkonen et al. (1988) was concerned with a linearity test against ESTAR under stationarity and suggested overcoming the problem of unidentified parameters by applying a Taylor approximation of G around $\gamma = 0$. The same procedure can be used for non-stationary models as well which leads to the auxiliary regression

$$\Delta y_t = \psi y_{t-1}^3 + u_t \tag{5.2}$$

with $\psi = \gamma\phi$ and u_t being a noise term depending on ε_t, ϕ and the remainder of the Taylor expansion. In this regression, the pair of hypotheses is now $H_0 : \psi = 0, H_1 : \psi < 0$. Kapetanios et al. (2003) suggest a Dickey-Fuller-type test for this hypothesis given by

$$t_{\text{KSS}} \equiv t_{\psi=0} = \frac{\widehat{\psi}}{\sqrt{\text{var}(\widehat{\psi})}} = \frac{\sum_{t=1}^{T} y_{t-1}^3 \Delta y_t}{\sqrt{\widehat{\sigma}^2 \sum_{t=1}^{T} y_{t-1}^6}}, \tag{5.3}$$

where $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(\Delta y_t - \widehat{\psi} y_{t-1}^3)^2$ is the usual estimator of the error variance. Deterministic components as a constant or a constant and a linear trend are removed in a first step, i.e. one applies the test to de-meaned or de-trended data. This means that the test is actually applied to the residuals of the regression ($\widehat{u}_t$)

$$y_t = \beta' d_t + u_t$$

with $d_t = 1$ in the case of de-meaning or $d_t = [1,\ t]'$ in the case of de-trending, instead of y_t directly. For details concerning this test such as the asymptotic limiting distribution and critical values see Kapetanios et al. (2003).

5.2.2 Least Squares Grid Search Test

Park and Shintani (2005) propose a different way of handling unidentified parameters in the ESTAR model when testing the unit root hypothesis. Instead of applying a Taylor approximation of G the authors propose a grid search applied to the unidentified parameter. In contrast to the previous test, the pair of hypotheses is now $H_0 : \phi = 0$ vs. $H_1 : \phi < 0$, therefore the smoothness parameter γ is unidentified under H_0. Park and Shintani (2005) suggest to estimate the following least squares regression

$$\Delta y_t = \phi y_{t-1}(1 - \exp\{-\gamma y_{t-1}^2\}) + \varepsilon_t$$

for a sequence of fixed values for the smoothness parameter γ, i.e. $\gamma \in \Gamma = (10^{-1}P_T, 10^3 P_T)$ with $P_T = \sqrt{\sum y_t^2/T}$. This means that the grid size and its bounds depend on the sample variation of the considered time series. Park and Shintani (2005) use the infimum of the random sequence of t-statistics to test the null hypothesis of a unit root,

$$t_{\text{PS}} \equiv \inf_{\gamma\in\Gamma} t_{\phi=0}\ .$$

The limiting distribution of t_{PS} depends on the parameter grid Γ and the assumed transition function G, which is the exponential one in our work. Note that the framework of Park and Shintani (2005) allows a lot more types of transition functions. Kapetanios et al. (2003) show that the de-meaning and de-trending of the data in a first step and applying the unit root test in the second step leads to a similar asymptotic distribution with the difference that the standard Brownian motion has to be replaced with a de-meaned or de-trended one, respectively. In our Monte Carlo study, we simulate the small sample distribution of t_{PS} with de-meaned and de-trended data and provide critical values.

5.3 Unit Root Test against Markov Switching

We consider a MSAR model which has similar properties to the ESTAR model discussed in the previous section. As the mean and the variance are constant in this ESTAR model, we restrict our attention to regime switching in the autoregressive parameters in the case of MSAR models, too. The main difference between the ESTAR and the MSAR model is, at least from a statistical viewpoint, the regime switching mechanism. While a regime switch is driven by past and therefore observable values of the process itself (y_{t-1}) for the ESTAR model, an unobservable stochastic Markov process, labeled as (s_t), is the driving force in Markov Switching models. As no unit root test against this specific MSAR model exists, we newly develop such a test. Our test statistic is similar to the one suggested in Caner and Hansen (2001) and the treatment of unidentified parameters follows Hansen (1996) and Garcia (1998). In particular, we consider the following model under the alternative hypothesis:

$$\Delta y_t = \phi(s_t) y_{t-1} + \varepsilon_t \ , \tag{5.4}$$

where the autoregressive parameter $\phi(s_t)$ depends on the unobservable first order two state Markov chain (s_t) that takes the values one or two. Furthermore, it is assumed that (s_t) is irreducible and aperiodic, i.e. it is characterized by the transition probability matrix

$$\Pi = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

with $p_{ii} = P(s_t = i | s_{t-1} = i)$ for $i = 1, 2$, being the probability that the state process is in state i in period t, given that it was in the same state in the previous period.

Francq and Zakoïan (2001) show that a necessary and sufficient condition for stationarity is given by the following two inequalities:

$$\begin{aligned} c_1 &= p_{11}(1+\phi(1))^2 + p_{22}(1+\phi(2))^2 + (1-p_{11}-p_{22})(1+\phi(1))^2(1+\phi(2))^2 < 1 , \\ c_2 &= p_{11}(1+\phi(1))^2 + p_{22}(1+\phi(2))^2 < 2 . \end{aligned}$$

The null hypothesis that the process contains a unit root is given by

$$H_0 : \phi(1) = \phi(2) = 0 ,$$

while the alternative hypothesis of stationarity is $H_1 : \phi(1) < 0$ or $\phi(2) < 0$. The alternative $H_1^{\cap} : \phi(1) < 0$ and $\phi(2) < 0$ is a special case of H_1. If the test has power against H_1 it will have power against $H_1^{\cap}$ as well, cf. Caner and Hansen (2001). A one-sided Wald test statistic for H_0 against H_1 can be constructed in the spirit of Caner and Hansen (2001):

$$R = 1\left(\widehat{\phi}(1) < 0\right) t^2_{\phi(1)=0} + 1\left(\widehat{\phi}(2) < 0\right) t^2_{\phi(2)=0} ,$$

where $t_{\phi(i)=0}$ denotes the conventional t-statistic for the null hypothesis that $\phi(i)$ equals zero. Parameters are estimated via maximum likelihood. Note that the transition probabilities p_{11} and p_{22} are unidentified under the null hypothesis. In order to tackle this problem, we follow Garcia (1998) and consider the supremum of a sequence of test statistics $R(p_{11}, p_{22})$ where the transition probabilities take values of a bounded grid $\Gamma = (0, 1) \times (0, 1)$, i.e.

$$R^* = \sup_{p_{11}, p_{22} \in \Gamma} R(p_{11}, p_{22}) .$$

Garcia (1998) considered among other variants a linearity test against a first-order MSAR process under stationarity under both H_0 and H_1. Unfortunately, the asymptotic distribution of Garcia's test is not invariant to the value of the autoregressive parameter although his simulation results (Garcia (1998), Table 3) show that this dependence is negligible. Such non-invariance problem does not appear for the unit root test against MSAR because the parameter of our autoregressive process under H_0 is fixed at one (random walk) and is therefore unable to vary. In our Monte Carlo study we provide critical values for the R^* statistic. Regarding the deterministic terms, we follow the procedure suggested by Kapetanios et al. (2003), see our section 5.2.1. This means that data is de-meaned or de-trended before the unit root test is applied in order to cope with non-zero means or linear trends.

5.4 Monte Carlo Study

5.4.1 General Approach

The following Monte Carlo study is about the empirical power of different unit root tests against ESTAR and MSAR models under situations that are realistic in practice when analyzing the persistence properties of real exchange rates. We also investigate whether these tests are suitable to discriminate between these two models.

In a related study, Choi and Moh (2007) consider the behavior of various unit root tests against different non-linear alternatives. Among these tests are the Park and Shintani and Kapetanios et al. test which are considered in this paper as well. Choi and Moh find that all unit root tests have power against various non-linear alternatives. Whether a test has power does not depend on the correct specification of the alternative but on how far the alternative is away from the null of a unit root. However, Choi and Moh consider idealized parameter constellations and therefore obtain a satisfying power for each test. They do not consider real world parameter constellations which are the focus of this paper. As mentioned before, even unit root tests, specially constructed to detect non-linear stationary processes which are close to a unit root, hardly ever reject the null hypothesis of a linear unit root when applied to real exchange rates and therefore hardly ever support the PPP hypothesis.

This Monte Carlo study answers the question whether this might be due to a lack of power of the developed tests under realistic situations rather than to a correct decision of the test by not rejecting the unit root hypothesis. We also consider the question whether these tests have power against other non-linear alternatives. To be specific, we consider whether unit root tests against ESTAR have also power against Markov Switching processes and whether our Markov Switching test has power against ESTAR. If they do not have power against the other non-linear alternative this would help to select the correct model.

In general, unit root tests have good power properties in Monte Carlo studies relying on parameter constellations which do not appear in the analysis of real exchange rates. It is quite common to simulate processes with $N(0,1)$ innovations, but we account for small standard deviations that are often found empirically, see Rapach and Wohar (2006). Another issue is that the location parameter c in ESTAR models is usually assumed to be equal to zero which is not correct in many practical situations either. Especially as the Kapetanios et al. (2003) test is strongly based on that assumption this causes significant power losses (see Kruse (2008)). In order to obtain realistic parameter settings, estimations are carried out using data from the International Financial Statistics database from 1973:02 to 1996:12 for the DEM/USD, FRF/USD, GBP/USD and JPY/USD real exchange rates as done in Rapach and Wohar (2006) for ESTAR models. Their reported estimates are very close to those reported in Taylor et al. (2001). Due to the fact that the estimation results for the DEM/USD and the FRF/USD are quite similar, we do not consider the latter currency in our study. Since Markov Switching models are neither considered in Rapach and Wohar (2006) nor in Taylor et al. (2001), we fit the Markov Switching model described in section 5.3 to the same data set in order to achieve the highest degree of comparability.

The exact parameter constellations are given in Table 5.1 for the three considered pairs of currencies (JPY/USD, DEM/USD, GBP/USD). In each case we use first–order autoregressive models. An application of standard diagnostic tests (not given here to save space but are available upon request) suggest that these models are correctly specified. Starting with the ESTAR specifications, we observe that the smoothness parameter γ takes quite different values ranging from 0.165 (JPY/USD) to 0.449 (GBP/USD). Note that it is difficult to distinguish an ESTAR process that exhibits a small value of γ from

Table 5.1: Parameter Estimation Results

DEM/USD	
ESTAR	$\gamma = 0.264$, $c = -0.007$, $\sigma = 0.035$
MSAR	$[\phi(1), \phi(2)] = [-0.074, 0.007]$, $[p_{11}, p_{22}] = [0.917, 0.945]$, $\sigma = 0.028$
	$c_1 = 0.995$, $c_2 = 1.744$
GBP/USD	
ESTAR	$\gamma = 0.449$, $c = 0.150$, $\sigma = 0.033$
MSAR	$[\phi(1), \phi(2)] = [-0.310, 0.028]$, $[p_{11}, p_{22}] = [0.300, 0.860]$, $\sigma = 0.030$
	$c_1 = 0.971$, $c_2 = 1.052$
JPY/USD	
ESTAR	$\gamma = 0.165$, $c = 0.515$, $\sigma = 0.033$
MSAR	$[\phi(1), \phi(2)] = [-0.233, 0.001]$ $[p_{11}, p_{22}] = [0.235, 0.953]$, $\sigma = 0.030$
	$c_1 = 0.982$, $c_2 = 1.093$

Remarks: Estimated parameter values for DEM/USD, GBP/USD and JPY/USD are taken from Rapach and Wohar (2006) for ESTAR models. Markov Switching models are estimated via conditional maximum likelihood in Gauss.

a unit root process as $\Delta y_t = \varepsilon_t$ for $\gamma \to 0$. Therefore, the expected power is low for the JPY/USD parameter constellation and somewhat higher for the GBP/USD parameters. However, one should also bear in mind that small changes of γ near zero do change the behavior of the process significantly. We expect to find clear differences in the behavior of the tests for the different parameter constellations. The location parameter c varies also across currencies, while the estimated standard deviation of the error term σ is very low and far away from unity for each currency. It should be mentioned that the location parameter c is significantly different from zero in each case although it seems to be rather small for some currencies.

For the Markov Switching processes, we always find one stable mean reverting regime and a second regime with an autoregressive parameter that is slightly above but very close to zero implicating a unit root or a mildly explosive regime. The process is still globally stationary for all pairs of currencies because the two conditions (c_1 and c_2 in Table 5.1) derived in Francq and Zakoïan (2001) are not violated, see Table 5.1. There-

Table 5.2: Small Sample Critical Values

de-meaning									
$T = 250$	t_{DF}	t_{KSS}	t_{PS}	R^*	$T = 500$	t_{DF}	t_{KSS}	t_{PS}	R^*
1%	-3.46	-3.46	-3.66	27.42	1%	-3.44	-3.51	-3.70	29.26
5%	-2.88	-2.91	-3.13	18.65	5%	-2.87	-2.94	-3.12	20.03
10%	-2.57	-2.63	-2.82	15.02	10%	-2.57	-2.67	-2.83	16.20
de-trending									
$T = 250$	t_{DF}	t_{KSS}	t_{PS}	R^*	$T = 500$	t_{DF}	t_{KSS}	t_{PS}	R^*
1%	-3.99	-3.99	-4.23	31.53	1%	-3.98	-4.01	-4.23	32.85
5%	-3.43	-3.49	-3.67	22.62	5%	-3.42	-3.40	-3.68	22.98
10%	-3.13	-3.12	-3.39	18.40	10%	-3.13	-3.12	-3.36	19.00

fore, the cyclical behavior of real exchange rates can be reproduced. However, for the DEM/USD exchange rate, the parameter of the stable regime is almost zero. The state probabilities are also close to one for this currency whereas they are well between zero and one for the other two currencies. Consequently, in view of this the estimated model for the DEM/USD exchange rate is close to a unit root which means that the expected power of the Markov Switching unit root test is low for this exchange rate. This can also be seen by considering the values for c_1 and c_2. They imply that we can expect that the Markov Switching test has higher power when the estimated model for the British Pound is considered instead of the one for the German Mark. The estimated standard deviation is similar to that of the ESTAR models and thus again far away from unity for each currency.

We simulate 2,000 replications of each process and apply them to the standard Dickey-Fuller unit root test (denoted by DF) as a benchmark test, the unit root versus ESTAR tests by Kapetanios et al. (denoted by KSS) and Park and Shintani (2005) (denoted by PS), and the Markov Switching test proposed in section 5.3. The power is considered at the 5% level by using size adjusted small sample critical values obtained from 20,000 replications for sample sizes of $T = 250$ and $T = 500$ which corresponds approximately to 20 and 40 years of monthly data, respectively. Size-adjusted critical values are reported

Table 5.3: Empirical Power, $T = 250$

de-meaning	t_{DF}	t_{KSS}	t_{PS}	R^*	de-trending	t_{DF}	t_{KSS}	t_{PS}	R^*
JPY–ESTAR	10.5	10.1	9.7	2.6	JPY–ESTAR	8.2	7.0	7.0	9.2
JPY–MSAR	7.6	10.3	9.3	39.5	JPY–MSAR	6.1	6.7	7.8	37.7
DEM–ESTAR	11.2	12.9	11.7	2.7	DEM–ESTAR	9.1	7.8	8.6	9.3
DEM–MSAR	12.0	8.9	8.0	16.6	DEM–MSAR	7.5	5.3	5.0	13.4
GBP–ESTAR	14.3	15.7	14.7	4.8	GBP–ESTAR	10.2	10.5	10.9	10.5
GBP–MSAR	20.1	35.8	38.8	87.1	GBP–MSAR	11.5	24.5	27.4	79.7

in Table 5.2. Note, that we simulate processes of length $T + 100$ and delete the first hundred observations in order to reduce the effect of the starting value. It should be mentioned here that we use simulated small sample critical values for all tests and not just for the Markov Switching test in order to obtain comparability of the results. The critical values for all tests are given in Table 5.2.

Thus, we consider the power of the tests under the correctly specified alternative. This is an ESTAR model for the Kapetanios et al. and the Park and Shintani test and a MSAR model for the Markov Switching test, as well as the power under a misspecified alternative which is the MSAR model for the ESTAR tests and the ESTAR model for the Markov Switching test. The alternative is misspecified for the standard Dickey-Fuller test for all considered models. The simulations for the ESTAR models were done in R whereas the Markov Switching part was simulated in Gauss.

5.4.2 Results

In this subsection, we discuss the power results for the non-linear unit root tests. Table 5.3 gives the power for a sample size of $T = 250$ observations. We consider all non-linear unit root tests after de-meaning as well as after de-trending as both deterministics can be reasonable for real exchange rate data. Note, that we include a constant or a constant and a linear trend term in the Dickey-Fuller test regression. However, it can be seen

Table 5.4: Empirical Power, $T = 500$

de-meaning	t_{DF}	t_{KSS}	t_{PS}	R^*	de-trending	t_{DF}	t_{KSS}	t_{PS}	R^*
JPY–ESTAR	16.1	18.8	18.3	2.0	JPY–ESTAR	11.4	10.8	11.2	20.1
JPY–MSAR	16.4	19.6	18.3	74.6	JPY–MSAR	9.3	11.2	11.4	70.9
DEM–ESTAR	22.3	29.7	29.7	2.2	DEM–ESTAR	13.4	14.2	15.8	21.7
DEM–MSAR	22.8	11.6	12.1	42.3	DEM–MSAR	13.1	7.5	8.6	30.8
GBP–ESTAR	30.5	49.1	50.9	1.9	GBP–ESTAR	22.4	23.9	26.0	23.7
GBP–MSAR	50.1	63.5	71.0	92.3	GBP–MSAR	31.6	49.1	56.5	93.9

that the results are rather similar in both cases. As we can see, neither of the ESTAR tests has considerable power against any of our models. Interestingly enough, the standard Dickey-Fuller test has higher power against ESTAR than the ESTAR tests for the JPY/USD and the DEM/USD in the de-trended case.

However, the power of all tests is extremely low when the true DGP is an ESTAR model in any case. This also holds for the Markov Switching test. When the true DGP is ESTAR, the Markov Switching test proves to be conservative. In opposition to the ESTAR tests, this is a rather convincing test property as a non-rejection of the test is the desired property for a correct model selection. Unfortunately, the ESTAR tests have power against the Markov Switching model. In each case, it is at least in the same region as the power against ESTAR models. For the GBP/USD it is far higher for the Markov Switching alternative than for the ESTAR alternative. Only for the DEM/USD exchange rate, the power of the tests is quite low. This was expected as the Markov Switching model is close to a unit root in this case. The Markov Switching test has satisfying power properties. Its power is quite high against a Markov Switching DGP except for the DEM/USD exchange rate where a low power was expected because of the near unit root structure of the DGP. On the other hand it has low power against ESTAR models. The DF test has similar power properties to the ESTAR tests.

Similar results can be observed for $T = 500$ (see Table 5.4). As expected, the power is generally higher compared to $T = 250$ but the results are qualitatively the same as before. The results for the de-trending case are qualitatively similar to those of the de-meaning case although all tests have less power under de-trending. This was expected as another deterministic parameter has to be fitted under de-trending. Unfortunately, the Markov Switching test is no longer conservative under de-trending when the true DGP is ESTAR. However, its power is still low and within the range of the ESTAR tests.

Altogether, we can say that there is no ESTAR test which dominates in terms of power. It can be argued, however, that all ESTAR tests have rather poor power against ESTAR with our parameter constellations which are realistic for real exchange rates. In some constellations the power of the ESTAR tests is even better for the Markov Switching alternative. As a result, by not rejecting the Null, these ESTAR tests do not allow us to conclude that the null hypothesis unit root is correct and therefore we cannot reject the Purchasing Power Parity hypothesis. However, when rejecting the Null we cannot conclude that the true model is ESTAR either. Further inference is necessary in order to select the correct model, but testing ESTAR against MSAR directly is quite complicated as it implies a non-nested testing problem. However, this issue is beyond the scope of this paper and left for future research.

5.4.3 Discussion

A natural question which arises out of this is why especially the ESTAR tests have so poor power properties. Figure 5.1 throws some light on this problem. In these graphs, the transition function of each estimated ESTAR process based on real data and parameters reported in Table 5.1 are depicted together with corresponding data points. Almost all data points are in the region where the transition function is close to its maximum. There are no data points at the tails of the function. Close to the maximum of the transition function the process behaves similarly to a unit root process or a highly persistent local-to-unity autoregressive process. The mean reverting property of the non-linear

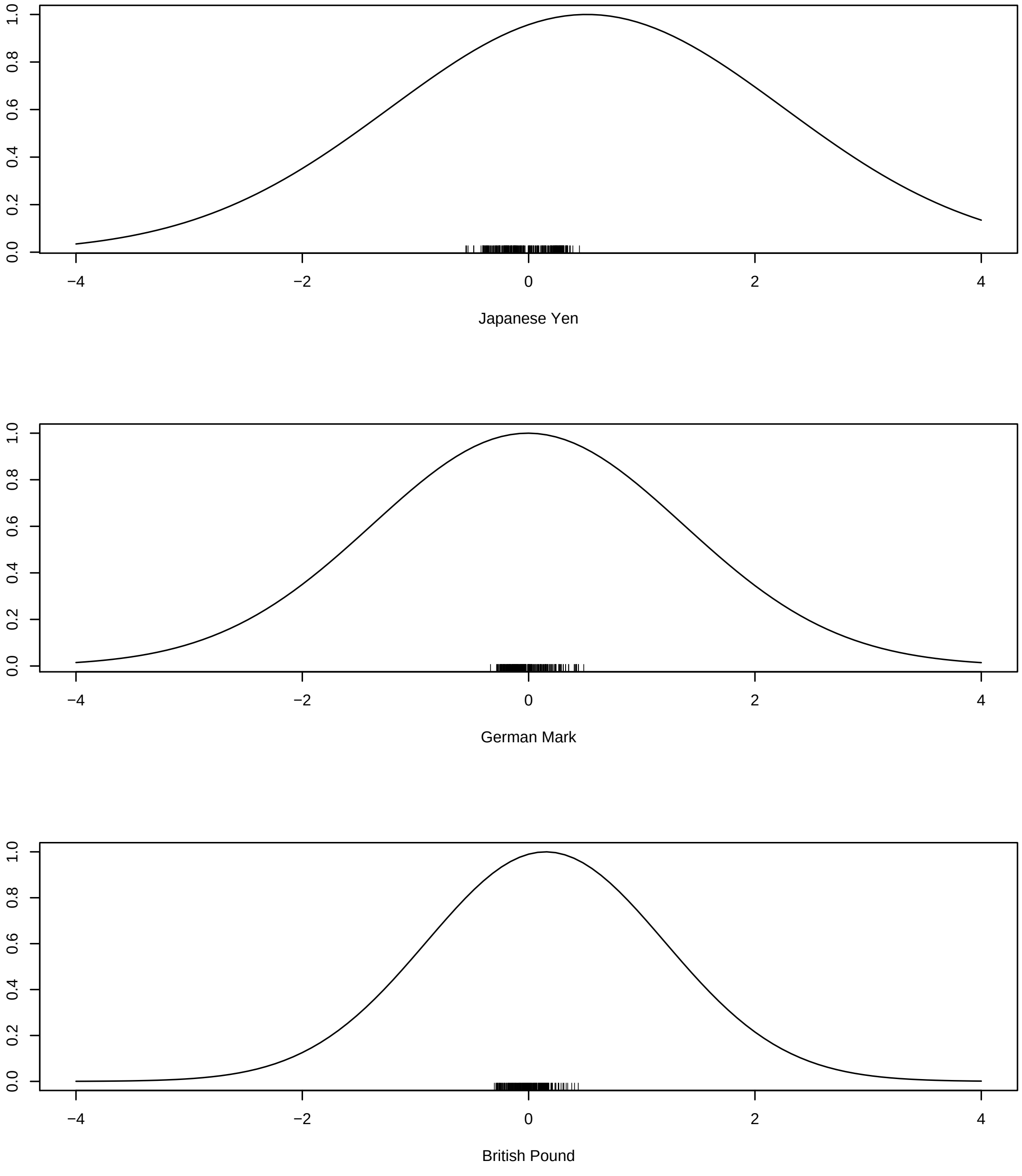


Figure 5.1: Estimated Transition Functions and Data Points.

time series model has a strong effect only in the outer regimes away from the equilibrium. Therefore, for the vast majorities of data points, the process behaves like a linear unit root process. This makes it hard or almost impossible for the tests to detect the non-linear mean reverting behavior of the DGP.

In addition to this, our simulation study shows that the ESTAR tests have similar power properties against ESTAR as against MSAR models. To intuitively explain this finding, we generate plots of y_{t-1} against the first difference $\Delta y_t = y_t - y_{t-1}$ for ESTAR and MSAR simulated time series generated from our parameter constellations and estimate the functional relationship between Δy_t and y_{t-1} in a non-parametric way by using the Naradaya-Watson estimator, see Figure 5.2. If the ESTAR effect is strong, the estimated curve should be near a cubic function. If the time series process has a unit root, it is identical to zero. As we can see, the cubic behavior is clearly pronounced for the DEM/USD and GBP/USD real exchange rate and less pronounced for the JPY/USD real exchange rate which is in line with our parameter settings. Moreover, it can be argued that both functions, the ESTAR and the MSAR function, are rather similar and quite close to each other. The MSAR process generates also a cubic shape for this function which is similar to the ESTAR model. As it can be argued that the idea of the Kapetanios et al. test is to check whether this function has a cubic trend or not, it detects the cubic form also for the MSAR process. As both functions are close to each other, the power is similar for both models.

This shows that the present tests are not able to detect ESTAR non-linearities as they are found in real exchange rates. Although the tests have convincing properties in many situations, they prove to have a lack of power under the very special parameter conditions which can be found in real exchange rates. It can be argued that it is rather difficult to draw any conclusion from the outcome of an ESTAR test under these conditions. Neither does a non-rejection of the Null mean that the true DGP which drive real exchange rates, is a linear unit root process nor does a rejection of Null mean that the true DGP is actually an ESTAR process.

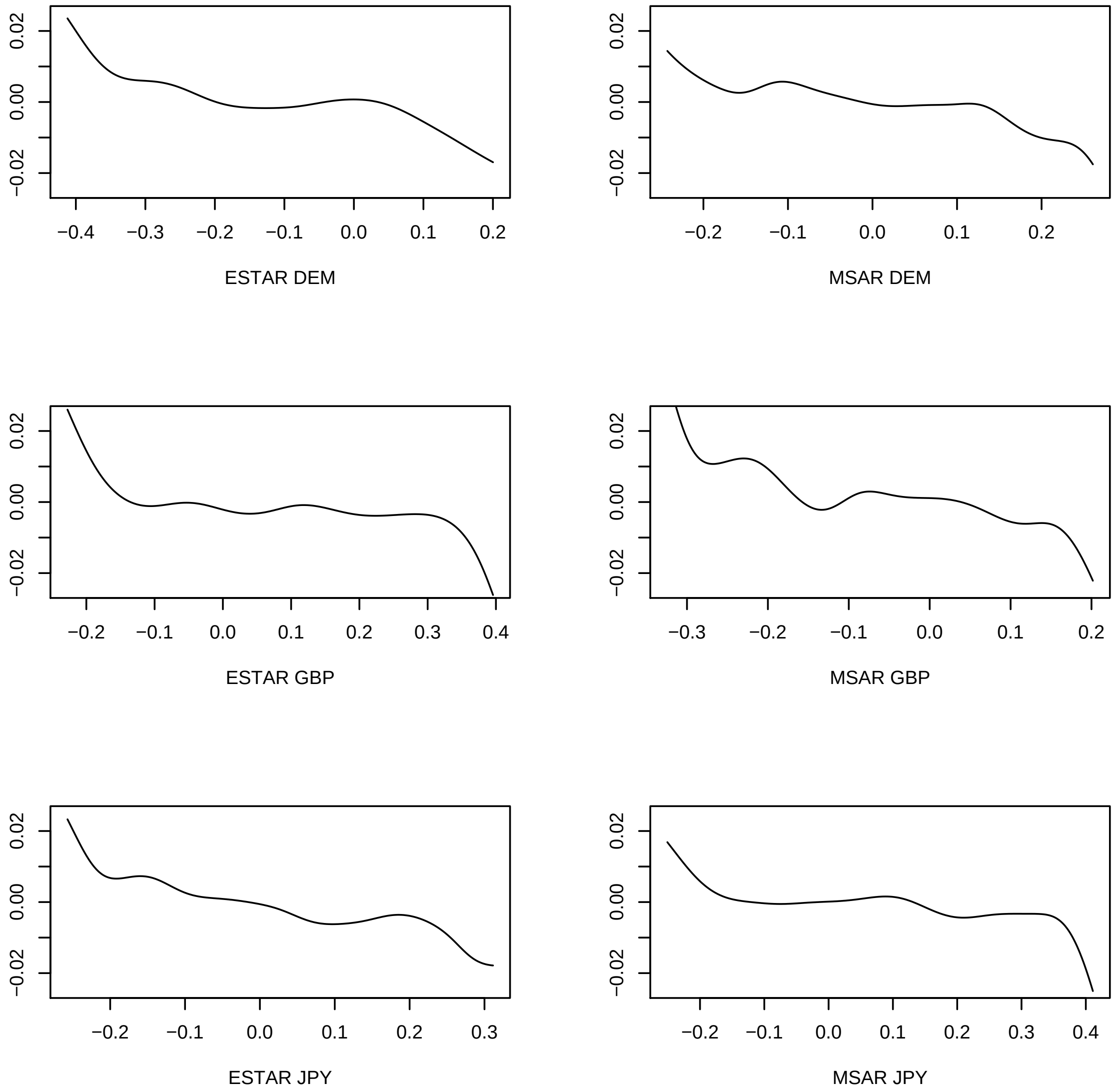


Figure 5.2: Nadaraya-Watson Estimates for the functional Relationship of ESTAR and MSAR Processes.

5.5 Application

This section applies the four unit root tests studied in the Monte Carlo simulations to the G7 exchange rates. Thus, we examine non-linearities in the real exchange rates of the US Dollar against the Canadian Dollar (CAD), Swiss Franc (CHF), German Mark (DEM), British Pound (GBP), Italian Lira (ITL) and Japanese Yen (JPY). Data is taken from the IMF International Financial Statistics database. Price levels are measured by the consumer price index (CPI). The sample covers the post-Bretton Woods period from 1973.01 to the Euro introduction 1998.12 implying a sample size of $T = 312$. This data set is chosen to achieve comparability to other studies and has the advantage that potential structural breaks that might have occurred due to the introduction of the Euro are excluded and thus not biasing our analysis. All time series seem to be persistent and locally trending, see Figure 5.3. The estimated partial autocorrelation functions (graphs are available upon request) indicate that all time series are first-order processes.

Next, we apply the standard Dickey-Fuller regression including a constant and test for linearity in the residuals. Linearity is tested by the neural network test proposed by Lee et al. (1993), Ramsey's RESET test (1969) and the BDS test for independence by Brock et al. (1996). These tests assume stationarity which is crucial when applied to real exchange rates themselves but not when applied to residuals. Results can be found in Table 5.5. They show that the linearity hypothesis has to be rejected in many cases. This also means that the Dickey-Fuller test regression neglects important non-linearities and is therefore misspecified. Recently, Harvey and Leybourne (2007) suggested a version of the classic linearity test against STAR models, originated by Luukkonen et al. (1988), that is robust against non-stationarity. However, such a robustification may reduce the test's power and we find only two rejections, namely for the German Mark and the Swiss Franc.

Moreover, we investigate the non-linearities by estimating the functional relationship between Δy_t and y_{t-1} in a non-parametric way by employing the Nadaraya-Watson estimator. Figure 5.4 shows these estimates. Only for the CAD/USD the estimated curve is

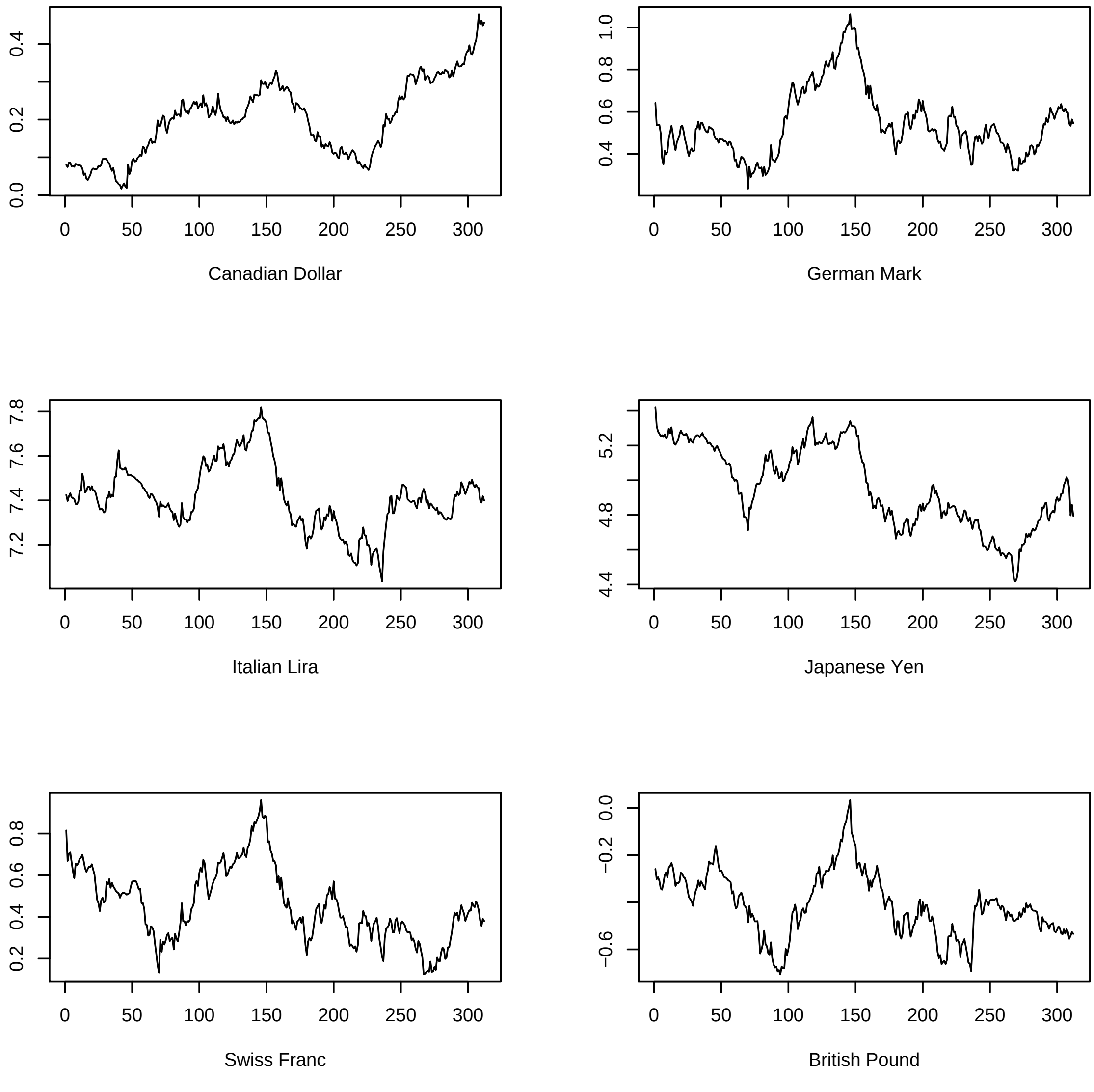


Figure 5.3: Logarithm of CPI-based Real Exchange Rates against US Dollar.

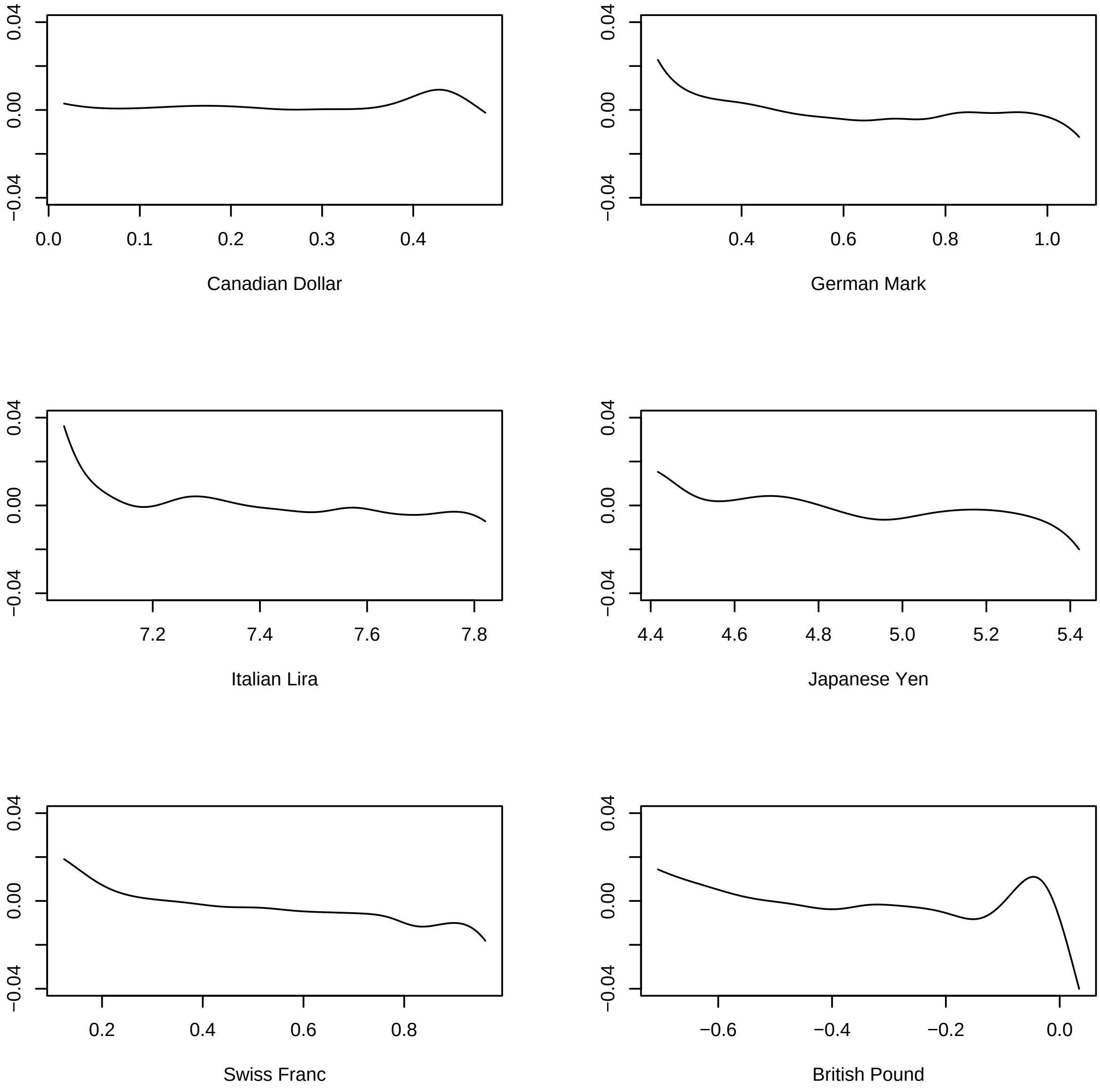


Figure 5.4: Nadaraya-Watson Estimates for the functional Relationship.

Table 5.5: Linearity and Unit Root Test Results

Test	CAD	CHF	DEM	GBP	ITL	JPY
			Linearity Tests			
NN	0.007	0.008	0.028	0.535	0.745	0.038
RESET(2)	0.885	0.204	0.118	0.185	0.243	0.038
RESET(3)	0.727	0.356	0.053	0.147	0.288	0.228
RESET(4)	NA	0.344	0.025	0.065	0.014	0.005
BDS(2)	0.135	0.006	0.054	0.008	0.012	0.408
BDS(3)	0.033	0.016	0.178	0.013	0.001	0.216
BDS(4)	0.045	0.007	0.270	0.009	0.000	0.210
HL	0.750	11.542	10.896	6.166	2.509	6.939
			Unit Root Tests			
t_{DF}	-0.10	-2.47	-1.92	-2.17	-1.84	-1.94
t_{KSS}	0.08	-2.54	-1.36	-2.46	-2.10	-2.45
t_{PS}	-0.12	-2.69	-1.95	-2.46	-2.09	-2.44
R^*	2.47	15.83	14.55	29.24	18.36	59.45

Notes: NN denotes the neural network test statistic by Lee et al. (1993). Hochberg's improved Bonferroni bound is used with one hundred draws to obtain reliable p-values for the neural network test, see Lee et al. (1993). RESET(m) is Ramsey's (1969) test statistic with terms up to power m+1. BDS(n) is the Brock et al. (1996) test statistic for independence with embedding dimension n. HL is the Harvey and Leybourne (2007) robust linearity test statistic calculated for $\alpha = 10\%$. For unit root tests, see Table 5.2.

very flat suggesting that there is no relationship between Δy_t and y_{t-1} which hints at a unit root. All other plots suggest the property of a mean-reversion and it is worthwhile to note that the functional relationship appears to be non-linear.

In the last step, we apply the four unit roots that have been studied previously in the Monte Carlo simulations to the six real exchange rate series in order to test empirically for the validity of PPP. Since all time series appear to be first-order processes, we do not include any lagged differences. The Dickey-Fuller regression contains a constant, while de-meaned data is used for all non-linear unit root tests.

The resulting test statistics are reported in the lower panel of Table 5.5. Neither the linear unit root test by Dickey and Fuller (1979) nor the non-linear unit root tests by Kapetanios et al. (2003) and Park and Shintani (2005) are able to reject the null hypothesis of a unit root at the ten percent level of significance. These results contradict the validity of PPP since there is no mean-reversion when a unit root is present. On the contrary, the new test against MSAR rejects the Null in favor of stationarity in four out of six cases. When having the outcomes of our preliminary analysis in mind, it is not surprising that the unit root hypothesis cannot be rejected in the case of CAD/USD. In addition, we note that the R^* statistic for the DEM/USD is quite close to the critical value of 15.02 which means that the test decision is borderline. Due to the fact that the Markov Switching unit root test does not have substantial power against ESTAR, especially in the case of de-meaned data, it is legitimate to conclude that there is no evidence for ESTAR dynamics in the data. Rather Markov Switching seems to be a more plausible model for explaining the dynamics of real exchange rates.

5.6 Conclusions

This paper provides a thorough empirical examination into the form of real exchange rate non-linearities. In particular, we investigate the power of unit root tests against ESTAR and Markov Switching and provide evidence supporting the relevance of Markov Switching processes in real exchange rates.

We contribute to the literature in four ways: Firstly, this research studies unit root tests under parameter settings that fit properties of real exchange rates. Secondly, we suggest a unit root test against Markov Switching autoregression that is similar in principle to the recently developed ESTAR tests and thus allows comparisons between both processes. Thirdly, we analyze the empirical power of these tests in an extensive Monte Carlo study where we consider a variety of ESTAR and MSAR processes. Finally, these tests are applied to the time series of the most important real exchange rates.

For each of these research directions we obtain findings that are of striking importance to Markov Switching processes. Referring to the first above mentioned direction, we find that the parameter setting is crucial for the power of ESTAR tests. Although these tests are powerful in general, under the specific conditions of currency markets, they seem to become clearly less useful. As we are interested in this kind of real exchange rate non-linearities, we need a unit root test against MSAR which is not available in the form we need here. Therefore, we propose a new test that builds upon inference techniques developed by Hansen (1996) and refined by Garcia (1998) as a second contribution. This brings us to the core of this research, which is to compare ESTAR and MSAR tests in a broad simulation study showing that ESTAR tests have poor power, whereas the MSAR test seems much more useful. Moreover, we observe that ESTAR tests are not robust with respect to Markov Switching dynamics while the opposite holds for our newly developed test. This means that a rejection of an ESTAR test, if any occurs, does not necessarily contain information about the type of non-linear adjustment to equilibrium. Finally, when applying these tests to important real exchange rates, we find that ESTAR tests cannot reject the unit root, whereas the MSAR test does this in most cases. This indicates that either PPP does not hold - which is not very plausible - or that ESTAR tests are not powerful - which seems to hold true - or that processes are not well described by ESTAR models, a possibility nourished by the finding of MSAR processes.

Overall, this research has an obvious economic implication that stems from the properties of MSAR vs. ESTAR processes. Whereas ESTAR models are used in international finance to capture the working of international arbitrage in goods and services, the Markov Switching model fits more with the idea of currency markets with heterogeneous agents whose interaction can create temporary exchange rate bubbles. This suggests that real exchange rate dynamics may be influenced to a substantial degree by destabilizing forces.

Chapter 6

Unit Roots and Smooth Transition Non-linearities

Co-authored with Philipp Sibbertsen

6.1 Introduction

Since the seminal work of Nelson and Plosser (1982) the question whether a time series contains a unit root and is therefore integrated of order 1, that is $I(1)$, or whether it is a globally stationary process, that is $I(0)$, attracted much attention in econometric research. As an $I(0)$ process can be interpreted as a process fluctuating around a stable equilibrium the question whether a given time series is $I(0)$ or $I(1)$ is equivalent to confirm or reject miscellaneous economic theories. Among many others, famous examples are the Purchasing Power Parity hypothesis or expectation hypothesis of the term structure. These theories are violated if a unit root is present in the real exchange rate or the term spread, respectively.

The literature on testing for a unit root against stationarity was for a long time concentrated on linear processes. However, in the last decade it became more and more clear that many economic time series are not linear. Various popular examples for highly non-linear processes among many others can be found in financial time series. Whereas early papers concentrated on the question whether a stationary process is linear or non-linear (see for example Luukkonen et al. (1988), Lee et al. (1993) or Brock et al. (1996)) the focus of econometric research recently is on testing for a unit root against globally stationary non-linear alternatives (see for example Kapetanios et al. (2003)). As most popular non-linear models can be interpreted as regime switching models in some sense they can have local unit roots although they are globally stationary and therefore, this problem is still challenging. However, these non-linear unit root tests concentrate on globally stationary alternatives which can be treated as $I(0)$ in the sense that the central limit theorem still holds. Some authors define an $I(0)$ process by the validity of the central limit theorem (see Davidson (2007)). Harvey and Leybourne (2007) also consider non-linear $I(1)$ processes with a global unit root.

Therefore, the practitioner has to distinguish between four different models. Each of these has a different economic interpretation and other implications in terms of forecasting, economic modeling and analysis of impulse-response functions. The aim of this paper is to give the empirically working econometrician a decision rule at hand which allows her a reliable classification of her time series into one of these model classes. As a measure of performance we use the number of correct classifications instead of size and power as we do not propose a correctly sized test in this paper. It should also be mentioned that we focus on STAR non-linearity which is one of the most popular non-linear models. As there are innumerable non-linear models around, it is impossible to create a unified procedure for all of them but our procedure can easily be generalized to Threshold AR or Markov Switching AR non-linearities.

In this paper we propose two different types of decision rules. First, we discuss a simultaneous procedure based on a linearity and a stationarity test which are independently computed. This procedure was proposed by Harvey and Leybourne (2007). They use a linearity test which is robust against non-stationarity of the time series. As this robustification can lead to power losses, we propose an alternative procedure which can be seen as a two-step or sequential procedure. In a first step, a unit root or a stationarity test is applied and the test result is used to choose the appropriate linearity test regression. If the test suggests that the time series is $I(0)$ the Wald-type linearity test is based on a test regression in levels and otherwise in first differences. As a stationarity test we use the test of Harris et al. (2003) or alternatively the non-linear unit root test of Kapetanios et al. (2003). However, it turns out that both procedures have better classification rates if the Harris et al. (2003) test is used. More importantly, the two-step procedures outperform simultaneous decision rules and especially the one proposed by Harvey and Leybourne (2007).

The paper is organized as follows. Section 6.2 introduces STAR models. Section 6.3 describes the linearity tests used and section 6.4 gives the unit root and stationarity tests. In section 6.5 several decision rules are proposed and section 6.6 contains a Monte Carlo study showing the classification rates of the various decision rules. Section 6.7 includes empirical applications to US government bond yields, the one-month interbank rate and the spread between them. Section 6.8 concludes. All Tables and Figures can be found in the Appendix of this chapter.

6.2 Non-linear STAR model

In the following section we briefly discuss the often applied first-order stationary STAR process and a non-stationary variant of it that has been studied by Harvey and Leybourne (2007).

Non-linear stationary STAR model

Consider the non-linear data generating process (DGP) for y_t with constant μ and let time be $t = 1, 2, \ldots, T$,

$$y_t = \mu + v_t \tag{6.1}$$

$$v_t = \phi v_{t-1} + \delta f(v_{t-1}, \theta) v_{t-1} + \varepsilon_t \tag{6.2}$$

$$\varepsilon_t \sim i.i.d.(0, \sigma^2)\ . \tag{6.3}$$

The error term ε_t is assumed to be a white noise process with mean zero and variance σ^2.

The autoregressive parameters are ϕ and δ. Non-linearity arises due to the presence of the smooth transition function f that depends on the two-dimensional parameter vector $\theta = (\gamma, c)'$, where $\gamma > 0$ determines the shape and $c \in \mathbb{R}$ the location of f. Common specifications for f are the exponential (f_E) and the logistic (f_L) smooth transition function

$$f_E(v_{t-1}, \theta) = 1 - \exp\{-\gamma(v_{t-1} - c)^2\} \tag{6.4}$$

$$f_L(v_{t-1}, \theta) = \frac{2}{1 + \exp\{-\gamma(v_{t-1} - c)\}} - 1 \, . \tag{6.5}$$

The main difference between them is the symmetry of f_E and the asymmetry of f_L with respect to $v_{t-1} - c$. It is implicitly assumed that the DGP is self-exciting, which means that a lag of the process itself, the first lag in our case, is the transition variable. A further assumption is that there are two regimes with a smooth transition between them. The first regime is characterized by $f = \{f_E, f_L\} = 0$ and the second by $f = 1$,

$$v_t = \phi v_{t-1} + \varepsilon_t \, , \quad f = 0 \tag{6.6}$$

$$v_t = (\phi + \delta) v_{t-1} + \varepsilon_t \, , \quad f = 1 \, . \tag{6.7}$$

Since this model comprises a linear AR process in each regime, we can measure local persistence by the sign and the magnitude of the autoregressive parameter in the respective regime, which is given by ϕ and $\phi + \delta$, respectively. In particular, local persistence changes smoothly from ϕ to $\phi + \delta$. Note that all other characteristics of the process, e.g. the variance of the error term ε_t, are not changing as a regime switch occurs.

If $f = f_E$, then y_t is globally stationary if $|\phi + \delta| < 1$, while $|\phi \pm \delta| < 1$ must be fulfilled in order to achieve global stationarity under $f = f_L$, see Harvey and Leybourne (2007). Further note, that a local unit root ($\phi = 1$) or even local explosiveness ($\phi > 1$) is permitted in the case of an exponential smooth transition function ($f = f_E$) while maintaining global stationarity of y_t. If $f = f_L$ is specified, such behavior is ruled out due to stronger restrictions for stationarity.

Non-linear non-stationary STAR model

Analogously to this DGP, Harvey and Leybourne (2007) consider an $I(1)$ version of it where non-linearity enters through first differences, i.e.

$$y_t = \mu + v_t \tag{6.8}$$

$$\Delta v_t = \phi \Delta v_{t-1} + \lambda f(\Delta v_{t-1}, \theta) \Delta v_{t-1} + \varepsilon_t \, . \tag{6.9}$$

In contrast to the previously discussed DGP, this one has an autoregressive lag structure of two and is globally non-stationary, which becomes more obvious after some rearrangements,

$$y_t = \mu + v_t \tag{6.10}$$

$$v_t = [1 + \phi + \lambda f(\Delta v_{t-1}, \theta)] \, v_{t-1} - [\phi + \lambda f(\Delta v_{t-1}, \theta)] \, v_{t-2} + \varepsilon_t \, . \tag{6.11}$$

The autoregressive parameters sum up to one, implying at least one unit root, regardless of the value of the smooth transition function f. This property still holds if the two extremes of f are considered,

$$v_t = (1 + \phi) v_{t-1} - \phi v_{t-2} + \varepsilon_t \, , \quad f = 0 \tag{6.12}$$

$$v_t = (1 + \phi + \lambda) v_{t-1} - (\phi + \lambda) v_{t-2} + \varepsilon_t \, , \quad f = 1 \, . \tag{6.13}$$

Suppose that $f = f_E$. If $\phi = 0$, then the lag structure changes from one to two as a regime shift occurs. Furthermore, if $f = 0$, then the process exhibits a unit root and if $f = 1$, then one root lies on the unit circle while the other one lies outside of it as long as $0 < \lambda < 1$ holds. Moreover, if $\phi = 1$ and $-1 < \lambda < 0$ then two unit roots are present under $f = 0$, while for $f = 1$ one root lies again on the unit circle and the second lies outside. A third case that is studied in the following is given by the setting $\phi = 1.5$ and $-1.5 < \lambda < -0.5$ which implies one unit root and one root inside the unit circle for $f = 0$, and one unit root and one root outside the unit circle for $f = 1$. Hence, the non-linear non-stationary exponential STAR model can have very different local persistence properties while it is globally non-stationary. Similarly to the stationary non-linear STAR model, such dynamics are not permitted if the logistic transition function ($f = f_L$) is assumed.

6.3 Testing Time Series Linearity

The non-linear DGP's that were presented in the previous section become linear under the constraint that the smoothness parameter equals zero, i.e. $\gamma = 0$. This holds true for the stationary as well as for the non-stationary DGP. Additionally, linearity can be achieved by setting δ (for the stationary DGP) or λ (for the non-stationary DGP) equal to zero. This means that if $H_0 : \gamma = 0$ is tested against $H_1 : \gamma > 0$, δ or λ appears to be a nuisance parameter under the null hypothesis. This circumstance is often referred to as the Davies problem, see Davies (1987). Luukkonen et al. (1988) suggested to overcome this problem of unidentified parameters under H_0 by applying a Taylor approximation to the smooth transition function f around $\gamma = 0$. This approach has been widely adopted and also applied by Harvey and Leybourne (2007) who employ a second-order expansion. In particular, when such an approximation is applied to the stationary and non-stationary DGP respectively, we get

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + \varepsilon_t \tag{6.14}$$

$$\Delta y_t = \beta_0 + \beta_4 \Delta y_{t-1} + \beta_5 (\Delta y_{t-1})^2 + \beta_6 (\Delta y_{t-1})^3 + \varepsilon_t \,. \tag{6.15}$$

The first equation is the Taylor approximation for the stationary $I(0)$ process and the second equation is the one for the corresponding non-stationary $I(1)$ process defined above. These auxiliary regression serve as the basis for testing linearity which is done by testing $H_0' : \beta_2 = \beta_3 = 0$ (for the stationary DGP) or $H_0' : \beta_5 = \beta_6 = 0$ (for the non-stationary DGP). Allowing for both degrees of integration simultaneously, $I(0)$ as well as $I(1)$, Harvey and Leybourne (2007) propose a hybrid test regression that incorporates terms from both individual test regressions, i.e.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + \beta_4 \Delta y_{t-1} + \beta_5 (\Delta y_{t-1})^2 + \beta_6 (\Delta y_{t-1})^3 + \varepsilon_t \,. \tag{6.16}$$

Now, the null hypothesis of linearity corresponds to four restrictions formulated as $H_0' : \beta_2 = \beta_3 = \beta_5 = \beta_6 = 0$, while the alternative of non-linearity can be written as H_1' : at least one of $\beta_2, \beta_3, \beta_5, \beta_6 \neq 0$. Harvey and Leybourne (2007) suggest the Wald statistic

$$W_T = \frac{RSS_0 - RSS_1}{RSS_1 / T} \,, \tag{6.17}$$

where RSS_i denotes the sum of squared residuals under H_i and T is the number of observations used in the test regression. Let us denote the Wald statistic (6.17) computed via (6.14) by W_T^0 and the one computed via (6.15) by W_T^1, where the exponent indicates the (implicitly) assumed degree of integration. Standard results imply that the limiting distribution of W_T^d is $\chi^2(4)$ if y_t is $I(d)$ with $d = \{0,1\}$. Harvey and Leybourne (2007) derive the non-standard distribution of W_T^0 if y_t is a linear random walk. Following the notation of Harvey and Leybourne (2007), $W_0(=\chi^2(4))$ denotes the limiting distribution of W_T^0 under $y_t \sim I(0)$ and W_1 denotes its limiting distribution under $y_t \sim I(1)$. In order to achieve the same limiting distribution under both degrees of integration, Harvey and Leybourne (2007) make use of Vogelsang's (1998) approach. Consider the transformed Wald test statistic

$$W_T^* = \exp\{-bH_T\}W_T \ , \tag{6.18}$$

where b is a non-zero constant and H_T is a statistic for testing $I(1)$ versus $I(0)$ with a pivotal limiting distribution under the null hypothesis. In addition, it is necessary that it converges to zero in probability under the alternative. Harvey and Leybourne (2007) set $H_T = |DF_T|^{-1}$, with DF_T being the Dickey–Fuller t–statistic obtained from

$$y_t \quad = \quad \pi_0 + \pi_1 y_{t-1} + \kappa \Delta y_{t-1} + \varepsilon_t \ .$$

In order to have the same critical values under both degrees of integration, that is $P(W_0 > c_\alpha) = P(\exp\{-bH\}W_1 > c_\alpha) = \alpha$, the constant b, which depends on the significance level α, has to be chosen accordingly. Harvey and Leybourne (2007) provide a response surface by fitting a seventh–order polynomial. Therefore, asymptotic critical values can be computed easily for any desired significance level α. This approaches, however, makes it impossible to use p-values because the test statistic W_T^* depends on the significance level.

The test regressions (6.14) and (6.15) can be used instead of (6.16) if the degree of integration is known, which is hardly the case in practice. Harvey and Leybourne (2007) show that the robust test has a good overall performance, but the price paid for robustification against non-stationarity can be high in terms of power. For example, if $T = 300$, the power loss that results from using W_T^* instead of W_T^1 can be up to twenty or nearly thirty percent for exponential or logistic processes, respectively.

6.4 Testing for and against Unit Roots

Unit Root Test

The unit root test we consider in this paper is the one proposed by Kapetanios et al. (2003). It builds upon a first-order Taylor approximation of a stationary exponential STAR model. The resulting test regression reads

$$\Delta y_t = \psi y_{t-1}^3 + \sum_{i=1}^{p-1} \rho_i \Delta y_{t-i} + \varepsilon_t$$

where the error term ε_t contains the Taylor approximation remainder that equals zero under the null hypothesis $H_0 : \psi = 0$. The alternative hypothesis is given by $H_1 : \psi < 0$ which ensures global stationary. The authors suggest a Dickey-Fuller-type t-statistic given by

$$t_T = \frac{\hat{\psi}}{\sqrt{\hat{\text{var}}(\hat{\psi})}} = \frac{\sum_{t=1}^T y_{t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^T y_{t-1}^6}}, \tag{6.19}$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T (\Delta y_t - \hat{\psi} y_{t-1}^3)^2$ is the usual estimator of the error variance. For reasons of comparability with the stationarity test described in section 4.2 and of empirical relevance we use de-meaned data $\tilde{y}_t \equiv y_t - \bar{y}$ where $\bar{y}$ denotes the mean of y_t.

Following the proof of consistency given in Kapetanios et al. (2003), one can verify that the test is consistent against stationary linear autoregressive processes as well. However, little is known about the small sample performance of this test if the true data generating process is actually non-stationary but non-linear as well. In section 6, we conduct the empirical size and power of the Kapetanios et al. (2003) test if data is generated by (non-stationary) exponential and logistic STAR models. In the following, the lag length p is set equal to two because the non-stationary STAR process is of order two.

Stationarity Test

Harris et al. (2003) propose a test for stationarity against a unit root that is based on sample autocovariances. Define $a_{t,k} = \tilde{y}_t \tilde{y}_{t-k}$, where $\tilde{y}_t$ denotes the deviation of y_t from its mean $\bar{y} \equiv \frac{1}{T}\sum_{t=1}^{T} y_t$. The test statistic is given by

$$S_T = \frac{1}{T^{1/2}} \frac{\sum_{t=k+1}^{T} a_{t,k}}{\hat{\omega}(a_{t,k})} \xrightarrow{d} N(0,1) \tag{6.20}$$

where $\hat{\omega}(a_{t,k})^2$ is the Bartlett kernel-based long run variance estimator of $a_{t,k}$. More specifically,

$$\hat{\omega}(a_{t,k})^2 = \hat{\gamma}_0(a_{t,k}) + 2\sum_{j=1}^{l}\left(1 - \frac{j}{l}\right)\hat{\gamma}_j(a_{t,k}) \tag{6.21}$$

$$\hat{\gamma}_j(a_{t,k}) = \frac{1}{T}\sum_{t=j+k+1}^{T} a_{t,k}a_{t-j,k} \tag{6.22}$$

The test rejects the null hypothesis of stationarity for large values of S_T. Since the simulation study in Harris et al. (2003) is somewhat limited, we extend their simulation analysis by considering the empirical power of S_T if the data generating process is (non-)stationary and non-linear, see section 6.

6.5 Decision Rules

The benchmark decision rule is the one used in Harvey and Leybourne (2007) that can be classified as a simultaneous procedure since it consists of two independently computed test statistics. These two statistics are the Harris et al. (2003) stationarity statistic S_T and the robust linearity statistic W_T^*. This decision rule is referred to as R_1. If, for example, both tests lead to a rejection (R) of their respective null hypotheses, we conclude that the process is non-linear $I(1)$. This procedure is depicted as follows, where NR stands for a non-rejection:

$$S_T, W_T^* \xrightarrow{\text{NR, NR}} \text{L–}I(0)$$

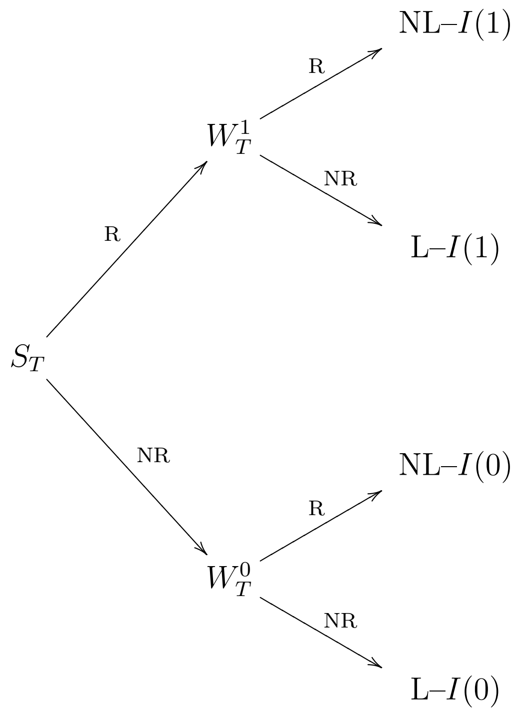
$$S_T, W_T^* \xrightarrow{\text{R, NR}} \text{L–}I(1)$$

$$S_T, W_T^* \xrightarrow{\text{NR, R}} \text{NL–}I(0)$$

$$S_T, W_T^* \xrightarrow{\text{R, R}} \text{NL–}I(1)$$

The success rate of a decision rule is measured as the the relative frequency of correct classifications. If, for example, the true DGP is linear $I(0)$, the success rate is simply the percentage of correct decisions for the category L–$I(0)$.

In the following section 6 we compare R_1 with a little modification of it, labeled as R_4, where we employ the Kapetanios et al. (2003) test instead of the Harris et al. (2003) test, by means of a Monte Carlo study. Moreover, we propose two versions of a two-step procedure that consists of a unit root or a stationarity statistic in a first step and a linearity test in the second step. The first-stage result is used to select the appropriate test regression: if the test in the first stage produces evidence for $I(0)$ we run the linearity test regression in levels, see equation (6.14), and otherwise in first differences, see equation (6.15). The decision rule R_2 applies the Kapetanios et al. (2003) test in the first step, while R_3 uses the Harris et al. (2003) test instead. Note, that the second step of R_2 and R_3 is identical. For example, the R_3 procedure can be depicted as:



The intuition behind the two-step procedures is as follows: the robustification of the linearity test that is achieved by applying W_T^* instead of W_T^0 or W_T^1 induces a substantial power loss which is not surprising at all. We analyze whether this power loss can be reduced by first selecting the appropriate linearity test regression given by equation (6.14) or (6.15). Note that the linearity test can be improved if the degree of integration is known, see Table 3 in Harvey and Leybourne (2007). In practice, the degree of integration is generally unknown, but if we can exploit the information of a test for $I(0)$ or $I(1)$ that exhibits enough power then the assumption of an unknown degree of integration becomes superfluous. Table 6.1 gives a short overview over different data generating processes and decision rules.

6.6 Monte Carlo Study

This section reports results of a variety of simulation experiments that shed light on the empirical small sample properties of the four decision rules. In a first step, we study the performance of the Kapetanios et al. (2003) unit root test and the Harris et al. (2003) stationarity test under non-linear $I(0)$ and $I(1)$ processes. On the one hand, it is not clear whether the Kapetanios et al. (2003) test is correctly sized under non-linear $I(1)$ DGP's and on the other hand there are no simulation results available yet that allow to draw conclusions about the behavior of the Harris et al. (2003) stationarity test under non-linear $I(0)$ processes. Furthermore, we are interested in the power of the Kapetanios et al. (2003) and the Harris et al. (2003) test if the DGP is of non-linear STAR-type. The interpretation of outcomes of different decision rules will be easier when we have these results in mind.

The parameter settings we consider are a large subset of those used in Harvey and Leybourne (2007) in order to achieve comparability. The nominal significance levels are one, five and ten percent for the size and power analysis of the Kapetanios et al. (2003) and the Harris et al. (2003) test, while we use a nominal five percent level of significance for the study of decision rules. The sample size is chosen as $T = 300$ which is also used in Harvey and Leybourne (2007). Following Harvey and Leybourne (2007), we set $k = [2T^{1/2}]$ and $l = [12(T/100)^{1/4}]$, where $[x]$ denotes the nearest integer of x, for the Harris et al. (2003) test. Moreover, the lag length (p) for the Kapetanios et al. (2003) test is set equal to one.

Table 6.2 reports the empirical sizes and powers of the Kapetanios et al. (2003) test and the Harris et al. (2003) test under ESTAR $I(0)/I(1)$ and LSTAR $I(0)/I(1)$ processes. We observe that the Kapetanios et al. (2003) test is a bit undersized if the the true DGP is non-linear $I(1)$. The only exception can be found in the last row. The same conclusions hold true for the Harris et al. (2003) test as well, where we also observe an exception. Although we do not provide an analytic proof, we conjecture that the distributions of t_T and S_T depend on the parameters of the non-linear DGP because their performance varies with these parameters. In order to cope with this problem a suitable bootstrap algorithm could be used to obtain more accurate critical values which is beyond the scope of this paper. Regarding the power properties both tests show quite good performance.

The Kapetanios et al. (2003) test often reaches hundred percent of power, while the power of the Harris et al. (2003) test is also quite high. In particular, both tests appear to be very powerful against logistic STAR models.

When analyzing the performance of the four decision rules we choose a nominal significance level of five percent. We start with linear first-order autoregressive processes, labeled as L–$I(0)$, see Table 6.3. The autoregressive parameter ϕ takes the values $0.00, 0.30, \ldots, 0.99, 1.00$. The local-to-unity values are chosen because it gets very difficult for unit root and stationarity tests to distinguish $I(0)$ from $I(1)$ processes in this region. If ϕ lies between zero and 0.7 we cannot observe big differences between the four decision rules and all of them show very good performance. In the local-to-unity region, the performance of R_2 and R_4, which are both based on the Kapetanios et al. (2003) test, worsens dramatically. On the contrary, the decision rules R_1 and R_3, which both use the Harris et al. (2003) test, are performing relatively good although the frequency of correct decisions is not extremely high at $\phi = 0.99$, but this could not be expected anyway. Wrong decision are made clearly in the direction of linear $I(1)$ processes which is due to the behavior of the unit root and the stationarity test.

Next, the performance of R_1 to R_4 is analyzed when the true DGP is a stationary non-linear exponential STAR process. Results are reported in Table 6.4. Both simultaneous rules (R_1 and R_4) are clearly outperformed by both two-step rules and in particular by R_2. However, in the experiments where $\phi = 0$, R_2 is dominated by R_3. Furthermore, the results for local unit root ($\phi = 1.0$) and local explosiveness ($\phi = 1.5$) suggest that the differences between both two-step methods are not big but R_2 dominates R_3. One exception to this is the case of $\phi = 1.5, \lambda = -1.0, \gamma = 0.1$, which is due to the fact that the Harris et al. (2003) test is definitely oversized in this case, see Table 6.2. The overall performance of R_2 and R_3 is satisfying and the gains with respect to R_1 range from four to nineteen percent. Additionally, we observe only little differences between the two simultaneous rules R_1 and R_4. Moreover, wrong decisions are often made in the direction of linear $I(0)$ processes because of the type II-error of linearity tests. It is rarely the case that the process is misclassified as non-stationary regardless of which decision rule is applied.

Turning to non-stationary exponential STAR processes, the results in Table 6.5 suggest that R_2 is best performing, followed by R_3. The gains from using a two-step procedure are evident as they range from four to twenty-six percent and they are higher on average than for ESTAR $I(0)$ processes. Again, R_1 and R_4 show quite similar performance but R_4 is preferable to R_1. Nonetheless, their success rates are far below those of R_2 and R_3. Our conclusions do not change a lot when interpreting the outcomes for logistic stationary (upper part of Table 6.6) and non-stationary (lower part of Table 6.6) STAR processes. Nonetheless, the differences between simultaneous and two-step rules are less pronounced for stationary processes. Furthermore, the frequency of correct decisions increases with the smoothness parameter γ in the case of logistic STAR processes because it does not become linear in limit ($\gamma \to \infty$). On the contrary, ESTAR processes become linear in the limit.

A unified two-step procedure

Recall that the two-step procedure based on the Kapetanios et al. (2003) test (R_2) performs relatively poor in the case of linear processes but very good for non-linear processes. Further note that the two-step procedure using the Harris et al. (2003) test (R_3) shows relatively good performance for linear processes as well. Hence, it is worthwhile to think of a procedure that takes the best out of both. One approach we suggest is to pre-test for linearity using the robust Wald statistic W_T^* and to proceed with R_2 in the case of a rejection and with R_3 in the case of a non-rejection. More formally, the unified decision rule, labeled as R_5 in the following, is defined by

$$\begin{aligned} \text{If } W_T^* &\geq \chi^2_{1-\alpha}(4), \text{ then } R_5 = R_2 \\ \text{If } W_T^* &< \chi^2_{1-\alpha}(4), \text{ then } R_5 = R_3 \, , \end{aligned}$$

where $\chi^2_{1-\alpha}(4)$ denotes the $(1-\alpha)\%$ asymptotic critical value for the W_T^* statistic. This unification is somehow in the spirit of the methodology used in Harvey et al. (2008b) which is based on the comments of Breitung to Harvey et al. (2008a).

Tables 6.7 and 6.8 report the results for the decision rule R_5 which clearly show that the proposed unification works very well. On the one hand, R_5 has the satisfying properties of R_3 when the true DGP is linear and on the other hand it shares the qualities of R_2 if the non-linearities are present. The power gains are obvious in the case of non-linear (non-)stationary STAR processes when compared to the performance of R_1, see Tables 6.4, 6.5, 6.6 and 6.8.

6.7 Empirical Application

The unified two-step procedure R_5 is applied to the US government bond yield, the one-month interbank rate and the spread between them. Data is taken from Datastream.[1] Our sample spans from 1986, February to May, 2008 and consists of 268 monthly observations. Figure 6.1 depicts the three time series. No clear trend can be detected in the spread by visual inspection and economic theory does not suggest that there are deterministic trends in interest rates, too. Therefore, we include only constants in the test regressions or we use de-meaned data.

As a by-product of this application we test the expectation hypothesis of the term structure (EHT) that requires the term spread to be stationary. However, the main aim of this empirical application is to classify the time series as (non-)linear and/or (non-)stationary. Such classification is of big importance for model building, the analysis of monetary shocks and for forecasting.

Results are reported in Table 6.9. As in Harvey and Leybourne (2007), we select the lag length for the test regressions by using a general-to-specific methodology at the ten

[1] The relevant codes are USGBOND. and BBUSD1M for the government bond yield and the interbank rate, respectively.

percent level of significance. The maximum lag order is set equal to four and the minimum equal to two. In a first step, we use the W_T^* test statistic in order to choose the appropriate decision rule which is R_2 for all time series, because W_T^* is significant in all cases. We conclude that the government bond yield and the interbank rate are non-stationary. In both cases, the Kapetanios et al. (2003) test does not reject the unit root hypothesis at the employed ten percent level of significance. On the contrary, the term spread appears to be stationary which supports the EHT and hints at cointegration between the government bond yield and the interbank rate. Although not reported in Table 6.7, the Harris et al. (2003) test confirms the conclusions drawn by the Kapetanios et al. (2003) test results, since S_T equals 2.721 (government bond yield), 1.702 (interbank rate) and 0.360 (term spread). The asymptotic critical value equals 1.282 at the nominal ten percent level of significance. Hence, the linearity test is carried out using first differences, see equation (6.15), in each case. In two cases we have to reject the null hypothesis of linearity in favor of STAR-type non-linearity. We conclude that non-linearities are more important for the shorter maturity and that the type of cointegration between the government bond yield and the interbank rate is in fact non-linear.

6.8 Conclusions

In this paper we propose a two-step decision rule based on a sequential procedure in order to classify a time series as either linear $I(0)$, linear $I(1)$, non-linear $I(0)$ or non-linear $I(1)$. The procedure is based on the subsequent application of a stationarity test by Harris et al. (2003) in a first step and, depending on the outcome of this test, the application of a Wald-type linearity test by Harvey and Leybourne (2007) to either the original time series or its first differences in a second step. In an extensive Monte Carlo study it is shown that this procedure and two variants of it outperform a simultaneous procedure suggested by Harvey and Leybourne (2007) in terms of better classification rates. Both approaches, the two-step as well as the simultaneous procedure are alternatively given with the non-linear unit root test of Kapetanios et al. (2003) or the stationarity test of Harris et al. (2003).

Whereas the procedure of Harvey and Leybourne (2007) consists of two simultaneously computed test statistics our proposed two-step decision rules make use of the outcome of a stationarity or unit root test in a first step before the appropriate linearity test is applied in a second step. Harvey and Leybourne (2007) use a linearity test which is robust against the degree of integration. Such a robustification may lead to power losses, especially in small samples. Therefore, the first stage tests' information about the degree of integration is used to apply the linearity test either to the original series if it appears to be $I(0)$ or its first differences if it it appears to be $I(1)$. It can be argued that the rates of correct classifications often increase significantly when sequential procedures are used. The gains are most pronounced for the case of an ESTAR-$I(1)$ process. The two-step procedures are also superior for the other cases in almost all situations.

As the two-step procedure has better classification rates for non-linear processes when the Kapetanios et al. (2003) test is used and worse rates with this test if the true DGP is linear it seems to be useful to pre-test the data by applying the robustified Wald-test of Harvey and Leybourne (2007). This unified procedure proves to give satisfying results in our Monte Carlo study and is therefore applied to the US government bond yield, the one-month interbank rate and the spread between them. We find that the bond yield is linear $I(1)$ whereas the interbank rate is non-linear $I(1)$ showing that non-linearities are more important for shorter maturities. The spread between both rates is classified as non-linear $I(0)$. On the one hand, this result confirms the EHT as the term spread can be treated as stationary and on the other hand, the cointegration relationship between the government bond yield and the interbank rate appears to be non-linear.

6.9 Appendix

Table 6.1: Data Generating Processes and Decision Rules

Type	Expression
L–$I(0)$	$(1-\phi L)y_t = \varepsilon_t$
L–$I(1)$	$(1-L)y_t = \varepsilon_t$
NL–$I(0)$	$(1-\phi L)y_t = \delta f(y_{t-1},\gamma)y_{t-1} + \varepsilon_t$
NL–$I(1)$	$(1-\phi L)\Delta y_t = \lambda f(\Delta y_{t-1},\gamma)\Delta y_{t-1} + \varepsilon_t$
R_1	Simultaneous, Harris et al. (2003) Test & W^*
R_2	Two-step, Kapetanios et al. (2003) Test & W_0/W_1
R_3	Two-step, Harris et al. (2003) Test & W_0/W_1
R_4	Simultaneous, Kapetanios et al. (2003) Test & W^*

Table 6.2: Empirical Size and Power of Kapetanios et al. (2003) and Harris et al. (2003) Test

			ESTAR–$I(1)$						ESTAR–$I(0)$					
			t_T			S_T			t_T			S_T		
ϕ	λ	γ	1.0	5.0	10.0	1.0	5.0	10.0	1.0	5.0	10.0	1.0	5.0	10.0
0.0	0.7	0.1	1.0	2.2	4.8	63.9	87.3	92.2	99.0	99.7	99.9	0.4	3.4	8.0
		0.5	1.4	3.5	4.9	63.2	87.0	92.2	84.1	95.2	97.8	0.4	3.4	8.8
		0.9	1.0	4.6	6.2	63.4	87.3	93.3	90.6	97.8	99.1	0.3	4.4	8.6
	0.9	0.1	1.0	5.0	9.1	63.7	86.7	92.4	94.6	97.5	98.2	0.4	3.6	8.5
		0.5	1.1	3.8	7.2	65.0	89.2	93.7	15.2	36.9	52.7	0.2	3.3	8.0
		0.9	1.0	4.2	8.0	68.6	91.0	94.9	26.2	55.0	70.2	0.4	4.8	10.3
1.0	-0.7	0.1	0.5	3.7	8.0	66.7	89.0	93.7	100.0	100.0	100.0	0.4	3.0	7.0
		0.5	0.6	3.9	8.7	63.2	87.3	92.8	100.0	100.0	100.0	0.5	4.1	9.0
		0.9	1.0	4.7	8.8	63.4	87.5	92.5	100.0	100.0	100.0	0.7	4.6	9.3
	-0.9	0.1	0.4	3.8	8.0	64.6	88.7	93.2	100.0	100.0	100.0	0.7	4.0	8.2
		0.5	0.6	4.4	9.1	62.9	88.1	92.8	100.0	100.0	100.0	0.6	4.1	8.1
		0.9	1.0	4.5	9.2	64.4	87.0	92.7	100.0	100.0	100.0	0.5	3.5	8.2
1.5	-1.0	0.1	0.0	0.4	1.1	87.6	98.4	99.1	90.6	96.0	97.7	9.2	27.6	40.0
		0.5	0.4	3.2	7.2	67.0	89.7	94.8	100.0	100.0	100.0	0.5	3.9	8.2
		0.9	0.8	4.4	9.2	64.5	87.6	92.5	100.0	100.0	100.0	0.6	3.6	8.2
	-1.4	0.1	0.1	1.0	3.2	76.6	94.4	96.8	100.0	100.0	100.0	1.0	6.1	12.6
		0.5	0.7	4.0	8.7	64.4	87.4	92.7	100.0	100.0	100.0	0.6	3.6	8.6
		0.9	1.2	5.0	9.6	65.7	88.6	93.3	100.0	100.0	100.0	0.4	4.0	9.4
			LSTAR–$I(1)$						LSTAR–$I(0)$					
0.0	0.7	0.1	0.7	4.0	8.4	65.1	88.5	93.4	100.0	100.0	100.0	0.2	3.6	8.2
		0.5	1.0	3.8	6.5	95.1	99.3	99.6	99.7	99.8	99.9	0.7	4.5	9.0
		0.9	2.3	5.3	7.9	99.6	99.9	100.0	99.3	99.8	99.9	0.5	4.1	8.8
	0.9	0.1	0.9	5.2	9.4	69.0	89.8	94.1	99.9	99.9	100.0	0.6	3.8	8.5
		0.5	1.8	4.8	7.3	98.6	99.9	100.0	98.5	99.2	99.6	0.4	3.7	9.1
		0.9	7.1	11.4	13.6	99.8	100.0	100.0	84.2	90.2	93.0	0.3	3.0	7.2

Table 6.3: Classification Rates for linear AR

ϕ	L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.00	89.6	94.5	93.7	94.9	4.8	0.0	1.6	0.0	5.3	5.5	4.6	5.1	0.3	0.0	0.1	0.0
0.30	92.5	95.4	91.4	96.5	3.3	0.0	3.7	0.0	3.9	4.6	4.9	3.5	0.3	0.0	0.0	0.0
0.50	92.8	96.9	91.8	96.1	3.8	0.0	3.9	0.0	3.3	3.1	4.0	3.9	0.1	0.0	0.3	0.0
0.70	92.7	94.4	90.2	94.5	3.6	0.3	4.8	0.1	3.5	5.3	4.8	5.4	0.2	0.0	0.2	0.0
0.90	88.6	74.8	87.2	75.2	6.7	18.6	7.5	19.3	3.9	5.4	4.8	4.7	0.8	1.2	0.5	0.8
0.95	72.1	37.6	71.4	33.0	23.0	55.6	22.5	59.3	3.3	5.0	5.5	5.0	1.6	1.8	0.6	2.7
0.99	21.0	5.5	23.2	6.9	73.3	87.6	70.9	87.1	0.3	2.0	1.8	1.2	5.4	4.9	4.1	4.8
1.00	12.7	3.1	9.1	3.6	82.6	91.4	85.8	90.4	0.9	0.6	1.6	1.2	3.8	4.9	3.5	4.8

Table 6.4: Classification Rates for non-linear ESTAR $I(0)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	68.2	61.6	60.7	72.7	2.8	0.2	3.1	0.0	28.0	38.2	36.1	26.8	1.0	0.0	0.1	0.5
		0.5	52.5	35.8	39.0	49.0	1.9	4.5	3.9	0.2	44.2	59.7	56.8	46.5	1.4	0.0	0.3	4.3
		0.9	77.5	76.2	73.4	81.9	2.9	1.7	2.5	0.5	18.6	22.1	24.0	15.7	1.0	0.0	0.1	1.9
	0.9	0.1	49.7	38.7	38.1	52.6	1.5	2.6	2.1	0.0	47.1	58.5	59.8	45.4	1.7	0.2	0.0	2.0
		0.5	45.8	17.2	29.7	24.4	1.9	62.8	3.3	27.8	50.9	17.1	66.8	11.3	1.4	2.9	0.2	36.5
		0.9	86.8	54.4	82.6	51.8	3.9	39.6	4.2	39.5	8.8	3.9	12.9	3.1	0.5	2.1	0.3	5.6
1.0	-0.7	0.1	20.1	9.2	9.1	17.5	0.3	0.0	3.7	0.0	76.7	90.8	86.8	82.5	2.9	0.0	0.4	0.0
		0.5	36.1	25.0	22.9	37.8	1.2	0.0	4.2	0.0	60.9	75.0	72.7	62.2	1.8	0.0	0.2	0.0
		0.9	64.3	54.6	52.2	66.9	1.8	0.0	3.1	0.0	32.1	45.4	44.4	33.1	1.8	0.0	0.3	0.0
	-0.9	0.1	7.0	3.1	3.3	9.6	0.1	0.0	4.6	0.0	89.9	96.9	90.9	90.4	3.0	0.0	1.2	0.0
		0.5	16.0	10.0	9.4	17.5	0.7	0.0	2.6	0.0	80.1	90.0	87.4	82.5	3.2	0.0	0.6	0.0
		0.9	43.0	33.2	32.0	44.6	1.6	0.0	2.8	0.0	52.7	66.8	65.0	55.4	2.7	0.0	0.2	0.0
1.5	-1.0	0.1	0.0	0.0	0.0	0.0	0.0	3.2	22.1	0.0	73.0	94.5	75.4	96.1	27.0	2.3	2.5	3.9
		0.5	6.5	2.8	2.0	6.8	0.2	0.0	4.8	0.0	90.1	97.2	92.3	93.2	3.2	0.0	0.9	0.0
		0.9	37.7	26.4	26.3	41.4	1.8	0.0	5.0	0.0	58.7	73.6	68.3	58.6	1.8	0.0	0.4	0.0
	-1.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	5.8	0.0	92.9	100.0	92.2	100.0	7.1	0.0	2.0	0.0
		0.5	0.2	0.1	0.0	0.1	0.0	0.0	3.1	0.0	96.0	100.0	95.4	99.9	3.8	0.0	1.5	0.0
		0.9	9.2	4.7	4.1	7.4	0.3	0.0	2.9	0.0	87.2	95.3	92.5	92.6	3.3	0.0	0.5	0.0

Table 6.5: Classification Rates for non-linear ESTAR $I(1)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	10.0	4.1	13.0	2.7	65.2	60.7	52.2	70.5	3.1	0.4	1.0	2.4	21.7	34.8	33.8	24.4
		0.5	7.5	4.2	10.9	1.3	49.3	35.2	34.2	55.7	5.0	0.6	0.6	2.5	38.2	60.0	54.3	40.5
		0.9	10.4	3.3	11.3	3.9	70.9	70.3	65.9	79.0	2.7	1.0	1.1	0.8	16.0	25.4	21.7	16.3
	0.9	0.1	6.5	4.9	12.7	2.4	50.8	39.4	35.5	53.1	7.2	0.4	0.9	3.1	35.5	55.3	50.9	41.4
		0.5	7.2	2.0	10.2	1.3	46.3	29.9	29.0	49.9	5.7	1.5	2.0	2.3	40.8	66.6	58.8	46.5
		0.9	7.2	2.6	4.7	3.5	81.4	81.4	81.6	85.1	0.7	1.1	1.8	0.6	10.7	14.9	11.9	10.8
1.0	-0.7	0.1	2.2	4.7	9.0	0.5	20.4	8.5	7.6	25.0	8.6	0.1	1.9	3.4	68.8	86.7	81.5	71.1
		0.5	4.0	3.8	12.7	1.7	38.8	23.8	22.5	43.9	7.9	0.1	0.3	3.4	49.3	72.3	64.4	51.0
		0.9	7.2	4.5	11.1	2.7	63.5	50.7	47.8	65.5	3.4	0.2	0.3	2.3	25.9	44.6	40.8	29.5
	-0.9	0.1	1.4	3.7	8.3	0.3	12.8	2.7	3.2	14.7	11.6	0.0	0.7	4.2	74.2	93.5	87.8	80.8
		0.5	3.4	4.2	12.6	0.8	20.2	9.2	7.6	24.4	11.1	0.0	0.5	5.2	65.3	86.6	79.3	69.6
		0.9	6.6	3.9	13.2	1.4	45.9	31.6	27.8	50.1	6.2	0.1	0.5	3.1	41.3	64.4	58.5	45.4
1.5	-1.0	0.1	0.0	0.5	1.3	0.0	4.2	0.0	0.0	4.1	0.8	0.0	0.9	0.3	95.0	99.5	97.8	95.6
		0.5	0.4	3.2	10.7	0.1	15.6	2.7	2.2	13.6	9.2	0.1	0.6	2.4	74.8	94.0	86.5	83.9
		0.9	4.6	3.7	10.7	1.8	40.6	31.8	23.7	45.7	7.0	0.1	0.8	2.5	47.8	64.4	64.8	50.0
	-1.4	0.1	0.0	0.5	5.5	0.0	3.4	0.0	0.0	2.3	5.5	0.0	1.1	0.9	91.1	99.5	93.4	96.8
		0.5	0.0	3.6	12.2	0.0	4.1	0.2	0.0	3.4	13.0	0.1	0.1	4.0	82.9	96.1	87.7	92.6
		0.9	1.0	4.4	11.2	0.3	13.6	3.5	3.2	15.7	11.9	0.1	0.5	4.5	73.5	92.0	85.1	79.5

Table 6.6: Classification Rates for non-linear LSTAR $I(0)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	88.9	89.5	88.2	90.9	3.6	0.0	2.6	0.0	7.2	10.5	8.7	9.1	0.3	0.0	0.5	0.0
		0.5	10.0	7.0	7.7	11.5	0.2	0.1	0.5	0.0	85.7	92.9	89.2	88.5	4.1	0.0	2.6	0.0
		0.9	0.2	0.0	0.1	0.6	0.0	0.1	0.6	0.0	95.6	99.4	96.7	99.4	4.2	0.5	2.6	0.0
	0.9	0.1	85.2	85.5	82.1	89.0	2.8	0.0	2.8	0.0	11.7	14.5	14.6	11.0	0.3	0.0	0.5	0.0
		0.5	1.4	0.5	1.0	2.5	0.1	0.1	1.3	0.0	94.8	99.0	95.8	97.2	3.7	0.4	1.9	0.3
		0.9	0.0	0.0	0.0	0.0	0.0	0.5	0.2	0.0	97.4	90.1	96.9	90.8	2.6	9.4	2.9	9.2

Table 6.6: Classification Rates for non-linear LSTAR $I(1)$

			L–$I(0)$				L–$I(1)$				NL–$I(0)$				NL–$I(1)$			
ϕ	λ	γ	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
0.0	0.7	0.1	8.3	2.8	9.0	2.9	82.9	86.1	79.4	88.2	0.8	0.8	1.3	0.7	8.0	10.3	10.3	8.2
		0.5	0.2	3.0	0.3	0.1	29.5	7.4	6.1	31.6	0.7	0.3	0.2	2.3	69.6	89.3	93.4	66.0
		0.9	0.0	2.1	0.0	0.1	12.2	0.0	0.1	14.0	0.0	0.6	0.0	2.6	87.8	97.2	99.9	83.3
	0.9	0.1	8.1	3.1	7.9	3.6	81.7	80.9	81.6	86.3	1.1	0.7	1.4	0.5	9.1	14.6	13.4	9.6
		0.5	0.0	2.3	0.2	0.0	19.4	0.4	0.2	16.8	0.2	0.4	0.0	3.3	80.4	97.1	99.0	79.9
		0.9	0.0	0.9	0.0	0.0	7.8	0.0	0.0	6.7	0.0	1.8	0.0	2.5	92.2	97.3	100.0	90.8

Table 6.7: Classification Rates of R_5 for linear AR

ϕ	L–$I(0)$	L–$I(1)$	NL–$I(0)$	NL–$I(1)$
0.00	90.6	4.5	4.9	0.0
0.30	92.4	2.6	5.0	0.0
0.50	92.0	3.5	4.5	0.0
0.70	92.0	3.6	4.4	0.0
0.90	86.7	8.6	4.3	0.4
0.95	70.1	24.2	4.6	1.1
0.99	21.8	72.6	1.4	4.2
1.00	11.2	83.2	0.9	4.7

Table 6.8: Classification Rates of R_5 for non-linear ESTAR $I(0)$ and $I(1)$

ϕ	λ	γ	L–$I(0)$	L–$I(1)$	NL–$I(0)$	NL–$I(1)$	L–$I(0)$	L–$I(1)$	NL–$I(0)$	NL–$I(1)$
0.0	0.7	0.1	60.0	2.5	37.5	0.0	11.0	54.6	0.5	33.9
		0.5	35.8	5.4	58.3	0.5	9.0	33.8	0.2	57.0
		0.9	73.0	5.2	21.6	0.2	12.3	65.7	0.0	22.0
	0.9	0.1	38.2	4.0	57.8	0.0	10.1	37.0	0.5	52.4
		0.5	30.6	47.7	19.5	2.2	5.8	29.0	2.1	63.1
		0.9	80.8	15.0	3.1	1.1	6.7	81.7	0.3	11.3
1.0	-0.7	0.1	8.7	0.4	90.9	0.0	4.3	8.7	0.0	87.0
		0.5	23.0	1.4	75.6	0.0	8.1	22.5	0.1	69.3
		0.9	50.8	2.1	47.1	0.0	10.9	47.0	0.0	42.1
	-0.9	0.1	3.9	0.3	95.8	0.0	4.1	3.3	0.0	92.6
		0.5	8.5	0.4	91.1	0.0	6.0	8.0	0.4	85.6
		0.9	31.1	1.6	67.3	0.0	7.9	31.5	0.5	60.1
1.5	-1.0	0.1	0.0	2.9	95.6	1.5	0.4	0.0	0.0	99.6
		0.5	2.4	0.3	97.3	0.0	4.0	2.4	0.2	93.4
		0.9	27.1	1.5	71.4	0.0	9.2	23.0	0.1	67.7
	-1.4	0.1	0.0	0.0	100.0	0.0	0.4	0.0	0.1	99.5
		0.5	0.0	0.0	100.0	0.0	4.0	0.1	0.2	95.7
		0.9	3.5	0.3	96.2	0.0	5.5	3.7	0.3	90.5
Classification Rates of R_5 for non-linear LSTAR $I(0)$ and $I(1)$										
0.0	0.7	0.1	88.5	3.2	8.3	0.0	10.0	78.8	0.6	10.6
		0.5	5.8	0.2	94.0	0.0	1.9	6.0	0.2	91.9
		0.9	0.1	0.1	99.6	0.2	1.8	0.0	0.9	97.3
	0.9	0.1	80.9	2.3	16.8	0.0	9.4	76.3	0.8	13.5
		0.5	0.6	0.1	98.8	0.5	3.2	0.6	0.6	95.6
		0.9	0.0	0.9	91.8	7.3	1.3	0.0	1.8	96.9

Table 6.9: Empirical Application to US Interest Rates

Time Series	W_T^*	t_T	S_T	W_T^0	W_T^1	Decision Rule R_5
Bond Yield	8.695*	−2.073	2.721*	—	3.458	Linear-$I(1)$
Interbank Rate	26.231*	−1.621	1.702*	—	31.740*	Non-linear-$I(1)$
Spread	27.181*	−3.027*	0.360	5.396*	—	Non-linear-$I(0)$

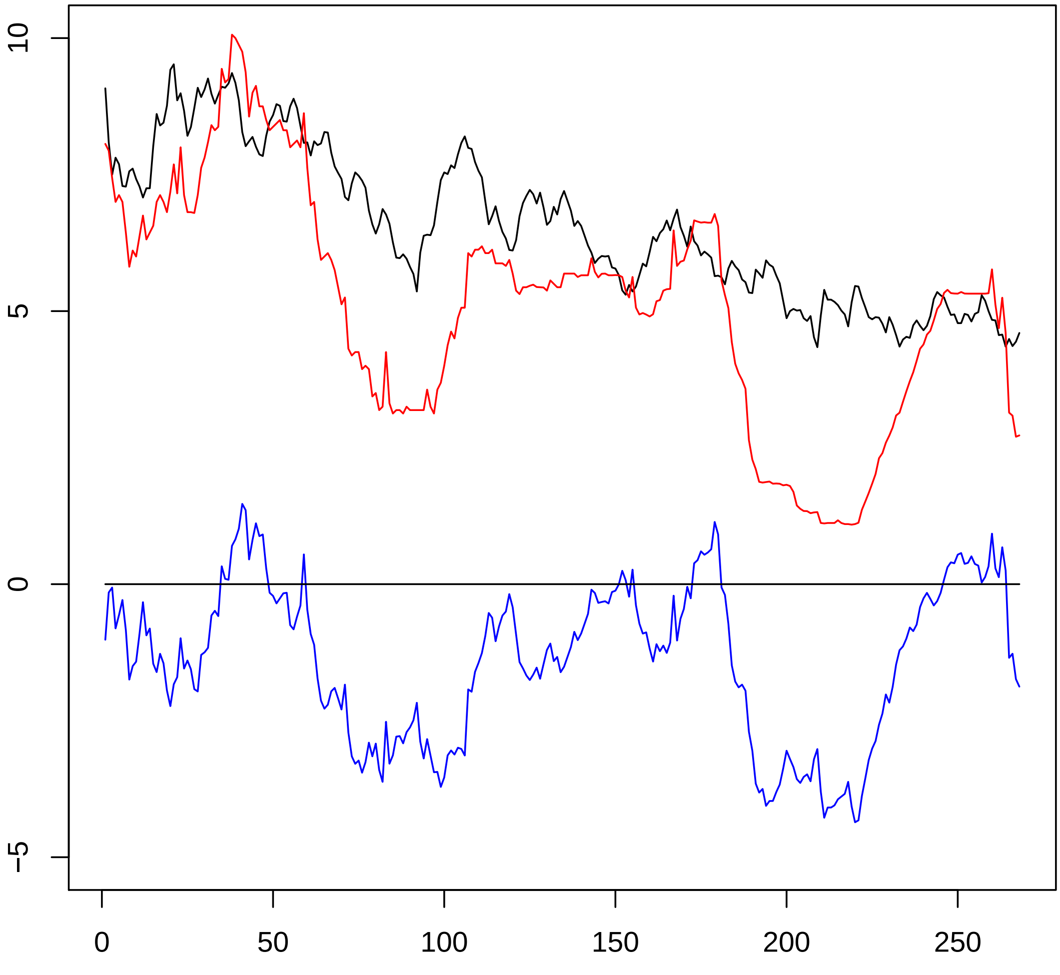


Figure 6.1: US Bond Yield (black), Interbank Rate (red) and Spread (blue).

Chapter 7

References

Abadir, K.M. and W. Distaso (2007): "Testing Joint Hypotheses when one of the Alternatives is One-Sided.", *Journal of Econometrics* 140, 695–718.

Banerjee, A., R. Lumsdaine and J. Stock (1992): "Recursive and Sequential Tests of the Unit Root and Trend Break Hypothesis: Theory and International Evidence." *Journal of Business and Economics Statistics* 10, 271–288.

Beran, J. (1995): "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models." *Journal of the Royal Statistical Society, Series B* 57, 659–672.

Beran, J. and N. Terrin (1996): "Testing for a Change of the Long-Memory Parameter." *Biometrika* 83, 627–638.

Brock, W.A., J.A. Scheinkman, W.D. Dechert and B. LeBaron (1996): "A Test for Independence Based on the Correlation Dimension." *Econometric Reviews* 15, 197–235.

Busetti, F. and A.M.R. Taylor (2004): "Tests of Stationarity against a Change in Persistence." *Journal of Econometrics* 123, 33–66.

Caner, M. and B.E. Hansen (2001): "Threshold Autoregression With a Unit Root." *Econometrica* 69, 1555–1596.

Cheung, Y.W. and U.G. Erlandsson (2004): "Exchange Rates and Markov Switching Dynamics." *CESifo Working Paper* No. 1348.

Choi, C.Y. and Y.K. Moh (2007): "How Useful are Tests for Unit-Root in Distinguishing Unit-Root Processes from Stationary but Non-Linear Processes?" *Econometrics Journal* 10, 82–112.

Chung, C.F. and R.T. Baillie (1993): "Small Sample Bias in Conditional Sum of-Squares Estimators of Fractionally Integrated ARMA Models." *Empirical Economics* 18, 791–806.

Davidson, J. (2007): "When is a Time Series I(0)?" Forthcoming in *The Methodology and Practice of Econometrics*, a Festschrift for David F. Hendry edited by Jennifer Castle and Neil Shepherd, Oxford University Press.

Davidson, J. and P. Sibbertsen (2007): "Tests of Bias in Log-Periodogram Regression." *mimeo.*

Davies, R.B. (1987): "Hypothesis Testing When a Nuisance Parameter is Present Under the Alternative." *Biometrika* 74, 33–43.

De Grauwe, P. and M. Grimaldi (2005): "Heterogeneity of Agents, Transaction Costs and the Exchange Rate." *Journal of Economic Dynamics and Control* 29, 691–719.

De Grauwe, P. and M. Grimaldi (2006): "Exchange Rate Puzzles: A Tale of Switching Attractors." *European Economic Review* 50, 1–33.

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.

Dolado, J.J., J. Gonzalo and L. Mayoral (2002): "Fractional Dickey-Fuller Test for Unit Roots." *Econometrica* 70, 1963–2006.

Doornik, J.A. and M. Ooms (2004): "Inference and Forecasting for ARFIMA Models With an Application to US and UK Inflation." *Studies in Nonlinear Dynamics & Econometrics* 8, Article 14.

Dumas, B. (1992): "Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World." *Review of Financial Studies* 5, 153–180.

Engel, C. (1994): "Can the Markov Switching Model Forecast Exchange Rates?" *Journal of International Economics* 36, 151–165.

Engel, C. and J.D. Hamilton (1990): "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?" *American Economic Review* 80, 689–713.

Francq, C. and J.M. Zakïoan (2001): "Stationarity of Multivariate Markov-Switching ARMA Models." *Journal of Econometrics* 102, 339–364.

Frankel, J.A. and K.A. Froot (1990): "Chartists, Fundamentalists, and Trading in the Foreign Exchange Market." *American Economic Review, Papers and Proceedings* 80, 181–185.

Frömmel, M., R. MacDonald and L. Menkhoff (2005): "Markov Switching Regimes in a Monetary Exchange Rate Model." *Economic Modelling* 22, 485–502.

Garcia, R. (1998): "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models." *International Economic Review* 39, 763–788.

Geweke, J. and S. Porter-Hudak (1983): "The Estimation and Application of Long-Memory Time Series Models." *Journal of Time Series Analysis* 4, 221–238.

Granger, C. and R. Joyeux (1980): "An Introduction to Long-Range Time Series Models and Fractional Differencing." *Journal of Time Series Analysis* 1, 15–30.

Hall, S.G., Z. Psaradakis and M. Sola (1997): "Switching Error Correction Models of House Prices in the United Kingdom." *Economic Modelling* 14, 517–527.

Hamilton, J.D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time-Series and the Business Cycle." *Econometrica* 57, 357–384.

Hamilton, J.D. (1994): "Time Series Analysis." Princeton University Press, Princeton.

Hansen, B.E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes." *Econometric Theory* 8, 489–500.

Hansen, B.E. (1996): "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis." *Econometrica* 64, 413–430.

Harris, D., B.P. McCabe and S.J. Leybourne (2003): "Some Limit Theory for Autocovariances Whose Order Depends on Sample Size." *Econometric Theory* 19, 829–864.

Harvey, D.I. and S.J. Leybourne (2007): "Testing for Time Series Linearity." *Econometrics Journal* 10, 149–165.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2008a): "Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition (with Commentaries and Rejoinder)."*Econometric Theory*, forthcoming.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2008b): "Testing for Unit Roots and the Impact of Quadratic Trends, with an Application to Relative Primary Commodity Prices." *Granger Centre Discussion Paper Series* 08/04.

Hassler, U. and J. Wolters (1995): "Long Memory in Inflation Rates: International Evidence." *Journal of Business and Economics Statistics* 13, 1326–1358.

Hurvich, C.M., R. Deo and J. Brodsky (1998): "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of a Long Memory Time Series." *Journal of Time Series Analysis* 19, 19–46.

Kanas, A. (2006): "Purchasing Power Parity and Markov Regime Switching." *Journal of Money, Credit, and Banking* 38, 1669–1687.

Kapetanios, G., Y. Shin and A. Snell (2003): "Testing for a Unit Root in the Non-Linear STAR Framework." *Journal of Econometrics* 112, 359–379.

Kang, K.H., C.J. Kim and J. Morley (2006): "Regime Shifts in U.S. Inflation Persistence." *mimeo.*

Kim, J. (2000): "Detection of Change in Persistence of a Linear Time Series." *Journal of Econometrics* 95, 97–116.

Kim, J., J. Belaire Franch and R. Badilli Amador (2002): "Corrigendum to Detection of Change in Persistence of a Linear Time Series." *Journal of Econometrics* 109, 389–392.

Koustas, Z. and A. Serletis (2005): "Rational Bubbles or Persistent Deviations from Market Fundamentals?" *Journal of Banking and Finance* 29, 2523–2539.

Kruse, R. (2008): "A New Unit Root Test against ESTAR Based on a Class of Modified Statistics." *mimeo.*

Lanne, M. (2006): "Nonlinear Dynamics of Interest Rate and Inflation." *Journal of Applied Econometrics* 21, 1157–1168.

Lee, T.H., H. White and C.W.J. Granger (1993): "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests." *Journal of Econometrics* 56, 269–290.

Leybourne, S.J., T. Kim, V. Smith and P. Newbold (2003): "Tests for a Change in Persistence against the Null of Difference Stationarity." *Econometrics Journal* 6, 291–311.

Leybourne, S.J., A.M.R. Taylor and T. Kim (2007): "CUSUM of Squares-Based Tests for a Change in Persistence." *Journal of Time Series Analysis* 28, 408–433.

Lothian, J.R. and M.P. Taylor (1996): "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries." *Journal of Political Economy* 104, 488–509.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988): "Testing Linearity against Smooth Transition Autoregressive Models." *Biometrika* 75, 491–499.

Michael, P., A.R. Nobay and D.A. Peel (1997): "Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation." *Journal of Political Economy* 105, 862–879.

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

O'Reilly, G. and K. Whelan (2005): "Has Euro Area Inflation Persistence Changed Over Time?" *The Review of Economics and Statistics* 87, 709–720.

Park, J.Y. and M. Shintani (2005): "Testing for a Unit Root against Transitional Autoregressive Models." *mimeo.*

Phillips, P.C.B. and P. Perron (1988): "Testing for a Unit Root in Time Series Regression." *Biometrika* 75, 335–346.

Phillips, P.C.B. and D. Sul (2003): "Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence." *Econometrics Journal* 6, 217–259.

Pivetta, F. and R. Reis (2007): "The Persistence of Inflation in the United States." *Journal of Economic Dynamics and Control* 31, 1326–1358.

Psaradakis, Z., M. Sola and F. Spagnolo (2004): "On Markov Error-Correction Models, with an Application to Stock Prices and Dividends." *Journal of Applied Econometrics* 19, 69–88.

R, Development Core Team (2004): "R: A Language and Environment for Statistical Computing." Available at www.r-project.org.

Ramsey, J.B. (1969): "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis." *Journal of the Royal Statistical Society. Series B* 31, 350–371.

Rapach, D.E. and M.E. Wohar (2006): "The Out-of-Sample Forecasting Performance of Nonlinear Models of Real Exchange Rate Behavior." *International Journal of Forecasting* 22, 341–361.

Rothe, C. and P. Sibbertsen (2006): "Phillips-Perron-Type Unit Root Tests in the Nonlinear ESTAR Framework." *Allgemeines Statistisches Archiv* 90, 439–456.

Sandberg, R. (2008): "Convergence to Stochastic Power Integrals for Dependent Heterogeneous Processes." *Econometric Theory*, forthcoming.

Sarantis, N. (1999): "Modeling Non-Linearities in Real Effective Exchange Rates." *Journal of International Money and Finance* 18, 27–45.

Sarno, L. (2005): "Towards a Solution to the Puzzles in Exchange Rate Economics: Where Do We Stand?" *Canadian Journal of Economics* 38, 673–708.

Sarno, L. and M.P. Taylor (2002): "Purchasing Power Parity and the Real Exchange Rate." *IMF Staff Papers* 49, 65–105.

Schwert, G.W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business and Economic Statistics* 7, 147–160.

Sercu, P., R. Uppal and C. Van Hulle (1995): "The Exchange Rate in the Presence of Transaction Costs: Implications for Tests of Purchasing Power Parity." *Journal of Finance* 50, 1309–1319.

Sibbertsen, P. and R. Kruse (2007): "Testing for a Break in Persistence under Long-Range Dependencies." *mimeo.*

Sollis, R. (2006): "Testing for Bubbles: An Application of Tests for Change in Persistence." *Applied Financial Economics* 16, 491–498.

Stock, J. and M. Watson (2003): "Has the Business Cycle Changed and Why?" In: M. Gertler and K. Rogoff, Editors, *NBER Macroeconomics Annual* 2002, MIT Press, Cambridge.

Taylor, A.M. and M.P. Taylor (2004): "The Purchasing Power Parity Debate." *Journal of Economic Perspectives* 18, 135–158.

Taylor, M.P. (2004): "Is Exchange Rate Intervention Effective?" *Economica* 71, 1–11.

Taylor, M.P. (2005): "Real Exchange Rates and Nonlinearities." in: P. de Grauwe, Editor, *Exchange Rate Economics: Where Do We Stand?*, MIT Press Cambridge Massachusetts, 87–123.

Taylor, M.P., D.A. Peel and L. Sarno (2001): "Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles." *International Economic Review* 42, 1015–1042 .

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

Vogelsang, T.J. (1998): "Trend Function Hypothesis Testing in the Presence of Serial Correlation." *Econometrica* 66, 123–148.