DISSERTATION

Econometric Models of Smooth Transition

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Ao. Univ. Prof. Dr. Bernhard Böhm
E105
Institut für Wirtschaftsmathematik

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Mag. Alenka Kavkler
Matrikelnummer: 0126921
Vrbanska cesta 12 b, 2000 Maribor, Slowenien

Wien, am 2. 2. 2006 ________________

Kurzfassung

Diese Dissertation widmet sich der Studie von ökonometrischen Modellen mit glatten Übergangsfunktionen (smooth transition regression models). Ein Überblick über solche Modelle nach dem letzten Stand der Literatur wird im zweiten und dritten Kapitel präsentiert. Verschiedene Formen der Funktionen, die den Übergang zwischen den Parameterregimes steuern, werden im zweiten Kapitel untersucht. Der dritte Abschnitt diskutiert die Frage von Spezifikation, Schätzung und Bewertung solcher Modelle. Für die verschiedenen Spezifikationstests, die bei der Suche nach einem passenden Regressionsmodell mit glatten Übergangsfunktionen erforderlich sind, werden verschiedene Varianten des Lagrange Multiplikator Tests verwendet und in einem Kapitelanhang erklärt.

Während ausführliche Untersuchungen auf dem Gebiet der univariaten nichtlinearen Modellbildung durchgeführt worden sind, muss die statistische Theorie der multivariaten nichtlinearen Modellbildung noch entwickelt werden. Diese Arbeit stellt in den weiteren Kapiteln dazu einige Beiträge zur Verfügung. Die ersten Versuche zur Erweiterung der nichtlinearen Regressionsmodelle mit glatten Übergangsfunktionen auf vektorautoregressive Modelle sind in den letzten Jahren erschienen. Ein Beitrag befasst sich mit dem Testen auf gemeinsame nichtlineare Komponenten in multiplen Zeitreihenmodellen. Spezifikation und Schätzung von Gleichungssystemen ist vereinfacht, weil die Existenz von gemeinsamen Nichtlinearitäten die Dimension der nichtlinearen Komponenten reduziert und damit eine sparsame Parametrisierung des Modells ermöglicht. Ein weiterer Beitrag betrifft Spezifikationsfragen in vektorautoregressiven Modellen, denen in einem weiteren Teil des vierten Kapitels nachgegangen wird. Die gegenwärtige Literatur über Spezifikations- und Schätzungsverfahren wird zusammengefasst und insofern erweitert als unterschiedliche Übergangsvariable und unterschiedliche Übergangsfunktionen in verschiedenen Gleichungen vorkommen. Die vorgeschlagene erweiterte Spezifikationsprozedur wird im letzten Teil des vierten Kapitels erklärt. Es wird gezeigt, dass die Zulassung verschiedener Typen von Nichtlinearitäten beziehungsweise unterschiedlicher Übergangsvariablen eine sinnvolle Verallgemeinerung ist.

Im fünften Kapitel präsentieren wir eine Anwendung des Tests auf gemeinsame Nichtlinearitäten um die Komponenten des realen Wechselkurses vom

Slowenischen Tolar gegenüber den Währungen von den fünf wichtigsten Handelspartnern Sloweniens zu analysieren. Bei nur einem der fünf untersuchten Fällen kann eine gemeinsame nichtlineare Komponente die Entwicklung der Komponenten des realen Wechselkurses ausreichend darstellen. Dieses Resultat entspricht der ökonomischen Realität und erlaubt eine sinnvolle Interpretation der geschichtlichen Entwicklung. Das sechste Kapitel widmet sich der Spezifikation eines nichtlinearen monetären Inflationsmodells. Dieses Modell besteht aus einer realen Geldnachfragegleichung erweitert um eine Phillips Kurve und dem Gesetz von Okun. Das Ziel dieser Untersuchung ist ein Vergleich der verschiedenen Eigenschaften, die durch die Spezifikation mit glatten Übergangsfunktionen auftreten, mit denen von linearen Modellen. Die Systemsimulation lässt auf eine für die Wirtschaftspolitik interessante Eigenschaft schliessen, dass sich nämlich monetäre Veränderungen auf Arbeitslosigkeit und Inflation asymmetrisch auswirken.

Abstract

This thesis is devoted to studying econometric models of smooth transition characterized by continuously changing parameter regimes. In chapter two a review of such models is presented. Various functional forms of the function that governs the transition between the regimes are examined. The following chapter discusses the issue of specifying, estimating and evaluating smooth transition models. The Lagrange multiplier test is usefully applied for the different specification tests required in the process of finding an appropriate smooth transition model. The review of the specification procedure, estimation and subsequent evaluation follows the recent literature.

Whereas there has been extensive research in the field of univariate nonlinear modeling, the statistical theory of multivariate nonlinear modeling has yet to be developed. The first attempts at extending nonlinear smooth transition regression techniques to vector autoregressive models have emerged in the last few years. One approach tries to test for common nonlinear components in multiple time series. Specification and estimation of the system of equations is simplified, since the existence of the common nonlinearities reduces the dimension of the nonlinear components in the system and enables parsimony. In a further section of chapter four a smooth transition approach is applied to vector autoregressive models. Starting from recent literature the specification and estimation techniques are reviewed and extended to the case of including more than one transition function and allowing for different transition variables in different equations. The proposed augmented specification procedure is explained in the last section of chapter four. It turns out that consideration of different types of nonlinearity or transition variables do indeed make sense.

In chapter five we present a practical implementation of the common nonlinearities test to analyze the components of the real exchange rates of the Slovenian Tolar versus the currencies of its five major trading partners. In only one of the five cases examined one common nonlinear component can adequately describe the development of the real exchange rate components. This result matches very well with economic reality and permits a good interpretation of the historic events. Chapter six is devoted to the specification of a nonlinear monetary model of inflation characterized by the real money demand equation augmented by the Phillips curve and the equation of Okun‘s

law. The objective of this investigation is to analyze different properties that derive from specification using smooth transition regression as compared to a linear specification. The simulation of the system reveals the fact interesting for the economic policy that a monetary impact leads to asymmetric effects on unemployment and inflation.

Acknowledgements

I would like to express my gratitude to my mentor, prof. dr. Bernhard Böhm, for his invaluable advice, kindness and encouragement throughout my PhD study. My thanks also go to my parents for their support and to Dušan for his patience.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Macroeconomic modeling has long been dominated by the assumption of linearity. Although it is generally excepted that the economy can be well described by a set of nonlinear relations among major economic variables, it was mainly because of simplicity that linear techniques to estimate such relations were usually applied in econometric model building. Individual equations were specified in order to maintain linearity in the estimated parameters, often using logarithmic or exponential transformations to achieve this property. With the progress of computing power it became easier to move from the spectrum of linear techniques to nonlinear ones. From the economic perspective, this move was particularly supported by the observation of serious structural changes in major economic relationships and the fact of asymmetric reactions to policy interventions.

The present investigation concentrates on a specific technique to analyze economic relationships, which exhibit certain structural properties. The properties are characterized by different parameter regimes that obtain under certain conditions. These are often mentioned by economic theory maintaining the idea that the economy behaves differently if values of certain variables lie in one region rather than in another, or, in other words, follow different regimes. A simple model that can generate such a situation would be the discrete switching model where parameters (or variables) visit a finite number of different regimes. The smooth transition regression model assumes that transitions between regimes are continuous. For many practical situations such a smooth transition seems to be a more realistic assumption that the abrupt switches. In chapter two a review of such models that have surfaced in the literature since the late 1980‘s is presented. Various functional forms of the function that governs the transition between the regimes are discussed. The following chapter is then devoted to the problem of specifying, estimating and evaluating smooth transition models. First, one has to start with the basic test of the null hypothesis of linearity. Since linearity is definitely the simplest assumption for a model builder to employ, it should be examined whether it is worth to deviate from it by considering the specific alternative of the nonlinear model. The Lagrange multiplier test will be usefully applied for the different specification tests required in the process of finding an appropriate smooth transition model. Because it appears in several versions a short survey is included as appendix to this chapter. The review of the specification procedure, estimation and subsequent evaluation follows the recent literature.

While for the estimation of single equation smooth transition regressions a well accepted specification procedure exists, this is quite different for the estimation of equation systems. As many issues in economics require the specification of several relationships, techniques to handle nonlinear features in systems are required. Only during the recent years such methods have appeared in the literature. Most of the work has been done in the framework of vector autoregressive models. Since every equation in a vector autoregressive model could also be estimated independently from the others (although not efficiently), such models lend themselves easily to an extension of the smooth transition approach. So far, two directions of research have appeared. One approach tries to test for common nonlinear components in multiple time series. The common nonlinearities approach is based on the canonical correlations technique. Specification and estimation of the system of equations is simplified, since the existence of the common nonlinearities reduces the dimension of the nonlinear components in the system and enables parsimony. The first section of chapter four critically reviews this approach. Applications of this technique are rare. In chapter five we present a practical implementation of this test devoted to an analysis of the components of the real exchange rates of the Slovenian Tolar versus the currencies of its five major trading partners. It turns out that in only one of the five cases examined one common nonlinear component can adequately describe the development of the real exchange rate components. This result matches very well with economic reality and permits a good interpretation of the historic events.

In the second section a smooth transition approach is applied to vector autoregressive models. Starting from recent literature the specification and estimation techniques are reviewed and extended to the case of including more than one transition function and allowing for different transition variables in different equations. The proposed augmented specification procedure is explained in the last section of chapter four. A model and the data are taken from Camacho [10] and subjected to a set of tests to discover possible transition variables and the types of the transition functions. It turns out that consideration of different types of nonlinearity or transition variables do indeed make sense. However, there is a greater ambiguity concerning the choice of transition variables in the different equations of the system.

Chapter six is devoted to the specification of a small system of equations based on the monetary approach to the explanation of inflation and unemployment. Data for Western Germany are used. The objective of this investigation is to analyze different properties that derive from specification using smooth transition regression as compared to a linear specification. Such studies are hardly to be found in the literature. Since the structure of the model is recursive, one can resort to an application of the single equation approach. The simulation of the system reveals the fact interesting for the economic policy that a monetary impact leads to asymmetric effects on unemployment and inflation.

From the economic point of view the current investigation indicates two aspects that could fruitfully be pursued in the construction of macro-models. The first one aims at the reduction of the dimension of nonlinear influences in order to achieve parsimonious models. Our application to the real exchange rate model shows that it is indeed possible to economize in this respect. The other aspect considers the variety of possible nonlinearities in the design of macroeconomic systems and their use for policy analysis. The potential to explore the extent of asymmetric effects should secure a relevant role for the smooth transition regression approach in econometric model building.

Chapter 2

Smooth transition regression models

2.1 Some simple nonlinear models

The building of macroeconomic models has long been dominated by the assumption of linearity. Nevertheless, it seems to be generally accepted that the economy is nonlinear, insofar as major economic variables are basically nonlinearly related.

Many elements of economic theory mention the idea that the economy behaves differently if values of certain variables lie in one region rather than in another, or, in other words, follow different regimes. The first attempt at modeling such phenomena is represented by the so-called discrete switching models, where a finite number of different regimes is assumed.

We start with a linear model of the form

$$y_t = x_t'\theta + \varepsilon_t, \tag{2.1}$$

where $x_t = (x_{t1}, x_{t2}, \ldots, x_{tk})'$ is a vector of explanatory variables of the model, $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ is a parameter vector and $\{\varepsilon_t\}$ are independent identically distributed errors with zero mean and constant variance. By assuming that the parameter vector θ is a function of a variable s_t we obtain a generalization of model (2.1):

$$y_t = x_t'\theta(s_t) + \varepsilon_t. \tag{2.2}$$

This locally linear approximation of the underlying nonlinear relationship between the observed variables can model an infinite number of different regimes. The simplest version takes into account only the two extreme regimes and can be written as

$$y_t = x_t'\theta_1 + (x_t'\theta_2) \cdot G(s_t) + \varepsilon_t, \tag{2.3}$$

where

$$G(s_t) = \begin{cases} 0, & \text{if } s_t < c \\ 1, & \text{if } s_t \geqslant c \end{cases}$$

is the *Heaviside function* indicating whether s_t exceeds the constant c or not. For obvious reasons, the model (2.3) is referred to as the *discreet switching regression model*.

The central tool of this class of models is the *switching variable* s_t that can be either observable or unobservable. However, the estimation of such models is complicated as the Heaviside function is not differentiable and the constant c is not known in advance.

Since the estimation of the discrete model (2.3) is not straightforward, it has been proposed by Goldfeld and Quandt [72] to replace the Heaviside function with the cumulative normal distribution function with mean c and constant variance σ^2, i.e.

$$F(s_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{s_t} \exp\left(-\frac{(s-c)^2}{2\sigma^2}\right) ds.$$

This can be interpreted as the first attempt to approximate the discrete step function (2.3) with a smooth function.

2.2 Types of smooth transition regression models

Due to the fact that smooth transition between regimes is often more convenient and realistic than just the sudden switches, several scientists proposed a generalization of discrete switching models of the following form.

Definition 2.1. *A model with a functional form*

$$y_t = x_t'\varphi + (x_t'\theta) \cdot G(\gamma, c; s_t) + u_t, \qquad t = 1, 2, \ldots, T, \tag{2.4}$$

where $x_t \in \mathbb{R}^{p+1}$ is the vector of explanatory variables containing lags of the endogenous variable y_t and the exogenous variables, i.e.

$$x_t = (1, x_{t1}, x_{t2}, \ldots, x_{tp})' = (1, y_{t-1}, \ldots, y_{t-m}, z_{t1}, \ldots, z_{tn})',$$

$\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_p)' \in \mathbb{R}^{p+1}$ and $\theta = (\theta_0, \theta_1, \ldots, \theta_p)' \in \mathbb{R}^{p+1}$ are the parameter vectors and u_t is a sequence of independent identically distributed errors, is a Smooth transition regression model (STR model).

G denotes a continuous *transition function* usually bounded between 0 and 1. Because of this property not only the two extreme states can be explained by the model, but also a continuum of states that lie between those two extremes. The *slope parameter* γ is an indicator of the speed of transition between 0 and 1, whereas the *threshold parameter* c points to where the transition takes place. The *transition variable* s_t is usually one of the explanatory variables or the time trend.

Let us explain in some detail the most often used transition functions.

2.2.1 LSTR1 model

The transition function of the *logistic STR model* or the *LSTR1 model* is defined as

$$G_1(\gamma, c; s_t) = \frac{1}{1 + e^{-\gamma(s_t - c)}}, \qquad \gamma > 0. \tag{2.5}$$

This type of the transition function was first proposed by Maddala [52], but it only became popular due to the work of professor Teräsvirta and coworkers (see [73] and [74]).

G_1 is a monotonously increasing function of the transition variable s_t, bounded between 0 and 1. Additionally, $G_1(\gamma, c; c) = 0.5$, therefore we can say that the location parameter c represents the point of transition between the two extreme regimes with $\lim_{s_t \to -\infty} G_1 = 0$ and $\lim_{s_t \to \infty} G_1 = 1$. The restriction $\gamma > 0$ is an identifying restriction. As we can see from Figure 2.1, the slope parameter γ indicates how rapidly the transition from 0 to 1 takes place. While a moderate value of $\gamma = 1$ imposes a slow transition, the function with $\gamma = 10$ changes quite fast.

If $\gamma \to \infty$ in the definition of G_1, then the model (2.4) converges to a switching regression model with the extreme regimes $y_t = x_t'\varphi + u_t$ and $y_t = x_t'(\varphi + \theta) + u_t$. For $\gamma = 0$, the function G_1 is constant and equal to 0.5. In this case model (2.4) simplifies to a linear regression model.

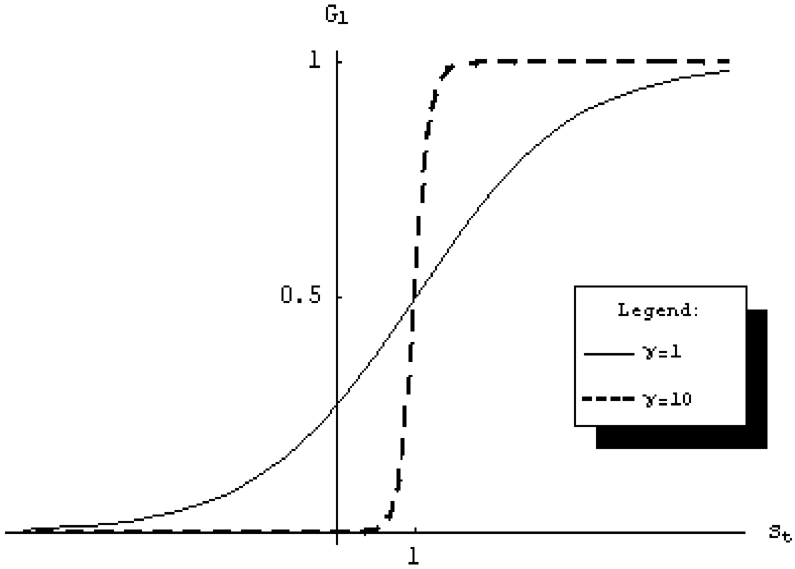

Figure 2.1: LSTR1 transition functions with $c = 1$

2.2.2 LSTR2 model

Monotonous transition may not always be satisfactory in applications. An example of a nonmonotonous transition function, which is especially useful in case of reswitching, is the quadratic logistic function

$$G_2(\gamma, c_1, c_2; s_t) = \frac{1}{1 + e^{-\gamma(s_t - c_1)(s_t - c_2)}}, \qquad \gamma > 0, \;\; c_1 \leq c_2. \tag{2.6}$$

Again, the restrictions on γ, c_1 and c_2 are identifying restrictions. G_2 is symmetric about the point $\frac{c_1+c_2}{2}$ and $\lim_{s_t \to \pm\infty} G_2(\gamma, c_1, c_2; s_t) = 1$. G_2 is never equal 0, its minimal value lies between 0 and 0.5. Two examples of the function G_2 with different values of the parameters are depicted in Figure 2.2.

As before, we consider the model obtained when $\gamma \to \infty$:

$$\lim_{\gamma \to \infty} G_2(\gamma, c_1, c_2; s_t) = \begin{cases} 1, & \text{if } \; s_t < c_1 \;\text{ or }\; s_t > c_2 \\ 0, & \text{if } \; c_1 < s_t < c_2 \\ \frac{1}{2}, & \text{if } \; s_t = c_1 \;\text{ or }\; s_t = c_2 \end{cases}$$

Setting $\gamma = 0$ again returns a linear model.

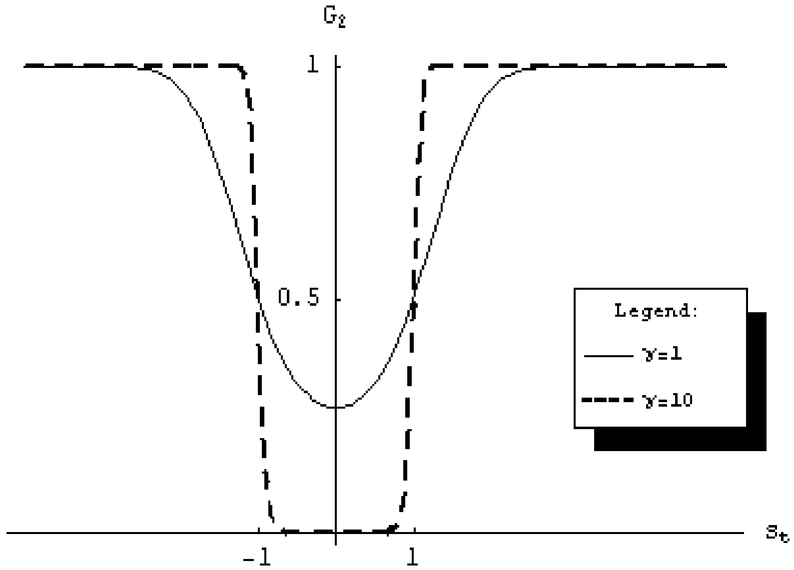

Figure 2.2: LSTR2 transition functions with $c_1 = -1$ and $c_2 = 1$

Due to the symmetric course of this transition function, the LSTR2 model is a special case of a three regime switching regression model that enables reswitching to the previous regime.

2.2.3 ESTR model

Sometimes it is desirable that small absolute values of the transition variable are related to small values of the transition function. The smooth transition regression model with an exponential transition function of the form

$$G_3(\gamma, c; s_t) = 1 - e^{-\gamma(s_t - c)^2}, \qquad \gamma > 0, \tag{2.7}$$

complies with the above condition for $c = 0$. This is the so-called *ESTR model*. The function G_3 is nonmonotonous and symmetric about the point c.

By setting $\gamma = 0$ we obtain a linear model.

$$\lim_{\gamma \to \infty} G_3(\gamma, c; s_t) = \begin{cases} 1, & \text{if } s_t \neq c \\ 0, & \text{if } s_t = c \end{cases}$$

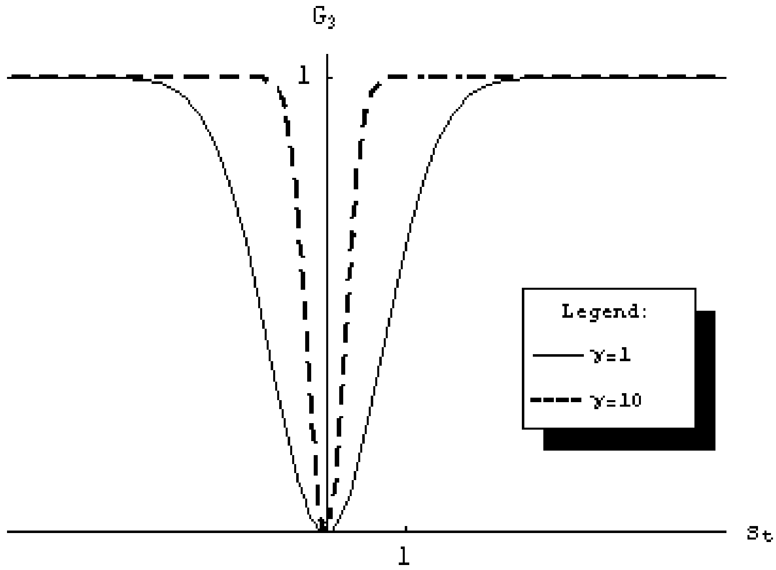

Figure 2.3: ESTR transition functions with $c = 0$

therefore it is in practice difficult to distinguish an ESTR model from a linear model for large values of γ.

Both the LSTR2 model and the ESTR model enable reswitching, but they differ in the rapidity of reswitching. One can see from Figure 2.3 that for a large value of γ the transition from 1 to 0 and back to 1 is sudden for the ESTR model, as compared to the LSTR2 model, where the reswitching can be slower when the gap between c_1 and c_2 is large.

2.2.4 Reswitching model

Two reswitches can be modeled with the nonmonotonous transition function introduced by Lin and Teräsvirta [48]:

$$G_4(\gamma, c_1, c_2, c_3; s_t) = \frac{1}{1 + e^{-\gamma(s_t - c_1)(s_t - c_2)(s_t - c_3)}}, \qquad \gamma > 0, \quad c_1 \leq c_2 \leq c_3. \tag{2.8}$$

The transition function (2.8) can be generalized to a case where the exponent is any polynomial of order 3 with real - valued roots. The Figure 2.4 depicts the transition function G_4 for $\gamma = 1$ and for $\gamma = 10$.

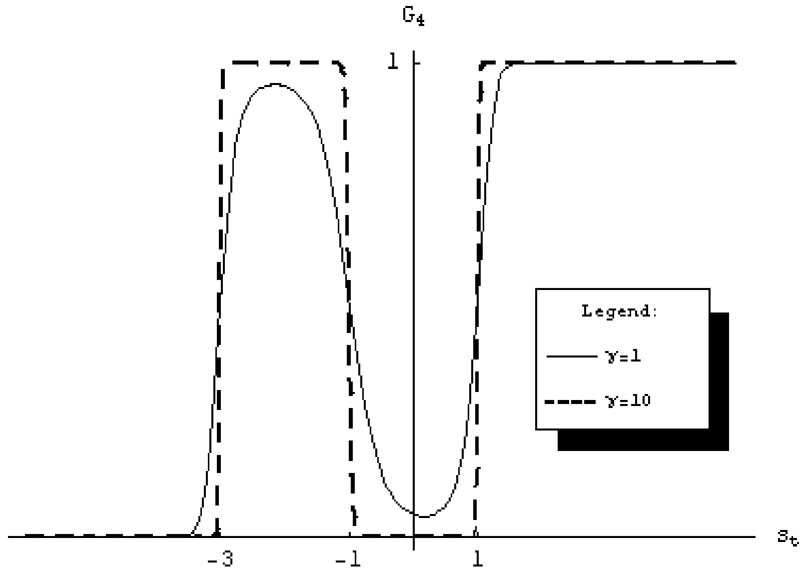

Figure 2.4: Reswitching functions with $c_1 = -3$, $c_2 = -1$ an $c_3 = 1$

As before, the γ - value determines the speed of transition between the extreme regimes. For $\gamma \to \infty$, we obtain a linear model:

$$\lim_{\gamma \to \infty} G_4(\gamma, c_1, c_2, c_3; s_t) = \begin{cases} 0, & \text{if } s_t < c_1 \text{ or } c_2 < s_t < c_3 \\ 1, & \text{if } c_1 < s_t < c_2 \text{ or } s_t > c_3 \\ \frac{1}{2}, & \text{if } s_t = c_1 \text{ or } s_t = c_2 \text{ or } s_t = c_3 \end{cases}$$

Chapter 3

Specification, estimation and evaluation of STR models

3.1 Testing linearity against STR

Before one starts the modeling cycle for a given data set, the following question should be cleared: Is the underlying relationship linear or nonlinear? The test that tries to answer this question was developed by Lukkonen, Saikkonen and Teräsvirta [51]. Their approach has two important advantages. Firstly, the asymptotic distribution under the null hypothesis of linearity is a standard distribution. Secondly, the test can be carried out just by using ordinary least squares. In order to discuss testing the statistical hypotheses within the STR framework, some additional assumptions about model (2.4) are necessary:

1. The stochastic variables among the exogenous variables $z_{t1}, z_{t2}, \ldots, z_{tn}$ are assumed nonstationary, whereas the nonstochastic ones are dummy variables.

2. s_t is assumed to be a stationary variable or the time trend.

3. Cross - moments $Ez_{ti}z_{tj}$, $Ez_{ti}s_t^k$, $Ey_{t-l}s_t^k$ and $Ey_{t-l}z_{tj}$ are assumed to exist for $k \leq 3$, $i, j = 1, 2, \ldots, n$ and $l = 1, 2, \ldots, m$.

We shall restrict our attention to LSTR1, LSTR2 and ESTR models. Next, we redefine the transition function as follows:

$$G_i^* = \begin{cases} G_i - \frac{1}{2}, & \text{for } i = 1, 2 \\ G_i, & \text{for } i = 3 \end{cases} \tag{3.1}$$

The advantage of G_i^* over G_i lies in the fact that G_i^* takes the value zero when $\gamma = 0$. Model (2.4) can be rewritten as

$$y_t = x_t'\varphi + (x_t'\theta) \cdot G_i^*(\gamma, c; s_t) + u_t, \qquad i = 1, 2, 3. \tag{3.2}$$

The parameter vectors φ and θ change, but the functional form stays the same.

Under additional assumption $u_t \sim N(0, \sigma^2)$, the conditional log - likelihood function of model (3.2) takes the form

$$\sum_{t=1}^{T} \ell(\varphi, \theta, \gamma, c; y_t | x_t, s_t) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} u_t^2. \tag{3.3}$$

The roots of the lag polynomial $1 - \sum_{j=1}^{m} \varphi_j L^j$ are assumed to lie outside the unit circle. The null hypothesis of linearity for model (3.2) can be expressed as

$$H_0 : \gamma = 0 \qquad \text{against} \qquad H_1 : \gamma > 0 \tag{3.4}$$

or as

$$H_0' : \theta = 0 \qquad \text{against} \qquad H_1' : \theta \neq 0. \tag{3.5}$$

This indicates an identification problem, because the model is identified under the alternative, but not identified under the null hypothesis. Namely, the parameters c and θ are nuisance parameters (for the null hypothesis $H_0 : \gamma = 0$) that are not present in the model under H_0 and whose values do not affect the value of the log - likelihood. Consequently, the Likelihood ratio test, the Lagrange multiplier and the Wald test do not have their standard asymptotic distributions under the null hypothesis and one cannot use these tests for a consistent estimation of the parameters c and θ.

To overcome this problem, Luukkonen, Saikkonen and Teräsvirta [51] replaced the transition function G_i^* with its Taylor approximation of a suitable order. Let us explain the procedure in more detail.

We start with the logistic transition function G_1^* and its first order Taylor approximation around $\gamma = 0$:

$$
\begin{aligned}
T_1(\gamma, c; s_t) &= G_1^*(0, c; s_t) + \left.\frac{\partial G_1^*(\gamma, c; s_t)}{\partial \gamma}\right|_{\gamma=0} \cdot \gamma + R_1(\gamma, c; s_t) \\
&= 0 + \left.\frac{(s_t - c) \cdot e^{-\gamma(s_t - c)}}{\left(1 + e^{-\gamma(s_t - c)}\right)^2}\right|_{\gamma=0} \cdot \gamma + R_1(\gamma, c; s_t) \\
&= \frac{1}{4}\,(s_t - c)\,\gamma + R_1(\gamma, c; s_t) \\
&= -\frac{1}{4}\,c\,\gamma + \frac{1}{4}\,\gamma\, s_t + R_1(\gamma, c; s_t) \\
&= \delta_0 + \delta_1\, s_t + R_1(\gamma, c; s_t), \qquad (3.6)
\end{aligned}
$$

where δ_0 and δ_1 are defined as $\delta_0 = -\frac{1}{4}\,c\,\gamma$ and $\delta_1 = \frac{1}{4}\,\gamma$. Replacing G_1^* with $T_1(\gamma, c; s_t)$ in (3.2) yields

$$
\begin{aligned}
y_t &= x_t'\varphi + (x_t'\theta) \cdot (\delta_0 + \delta_1\, s_t + R_1(\gamma, c; s_t)) + u_t \\
&= x_t'(\varphi + \delta_0\theta) + x_t's_t(\delta_1\theta) + x_t'\theta R_1(\gamma, c; s_t) + u_t \\
&= x_t'\eta_0 + x_t's_t\eta_1 + u_t^*, \qquad (3.7)
\end{aligned}
$$

with $u_t^* = x_t'\theta R_1(\gamma, c; s_t) + u_t$, $\eta_0 = \varphi + \delta_0\theta$ and $\eta_1 = \delta_1\theta$. Thus, the former null hypothesis of linearity $H_0 : \gamma = 0$ can also be tested as

$$
H_0'' : \eta_1 = 0 \qquad \text{against} \qquad H_1'' : \eta_1 \neq 0 \qquad (3.8)
$$

by performing a Lagrange multiplier (LM) type test. This test has the advantage that the estimation of the nonlinear model under the alternative hypothesis is not necessary. Different versions of the Lagrange multiplier test are discussed in the appendix to this chapter. Note that $u_t^* = u_t$ under H_0. Using the formula (3.101), the LM test statistic can be computed as

$$
LM = \frac{1}{\hat{\sigma}_R^2}\left(\sum_{t=1}^{T} w_t\hat{u}_t\right)'\left(M_{22} - M_{21}M_{11}^{-1}M_{12}\right)^{-1}\left(\sum_{t=1}^{T} w_t\hat{u}_t\right), \qquad (3.9)
$$

where

$$
z_t = x_t,\ w_t = x_t s_t,
$$

$$
M_{11} = \sum_{t=1}^{T} z_t z_t',\ M_{22} = \sum_{t=1}^{T} w_t w_t',\ M_{12} = M_{21}' = \sum_{t=1}^{T} z_t w_t',
$$

$\hat{u}_t$ is the residual estimated under the null hypothesis and the restricted estimator $\hat{\sigma}_R^2$ is equal to $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2$. The derivation of the above test statistic is described in detail in section 3.6.4. Under the null hypothesis, the statistic has an asymptotic χ^2 distribution with p+1 degrees of freedom, when the so-called regularity conditions are fulfilled. Basically, the moments used in the LM test statistic (3.9) have to exist (see section A.2.4 for details).

We have to emphasize that auxiliary regression (3.7) is suitable only if s_t does not belong to x_t. In the opposite case the variable s_t appears twice on the right-hand side of equation (3.7). The problem is solved by dropping the constant from the vector x_t by defining

$$\tilde{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tp})' \tag{3.10}$$

and transforming equation (3.7) into

$$y_t = x_t'\eta_0 + (\tilde{x}_t' s_t)\eta_1 + u_t^*. \tag{3.11}$$

The degree of freedom of the asymptotic χ^2 distribution is reduced by 1, since there are p restrictions left under the null hypothesis $H_0'' : \eta_1 = 0$.

Luukkonen, Saikkonen and Teräsvirta [50] investigated the power properties of the proposed linearity test by simulation. It turned out that the LM-type test has good power already in small samples, with the exception of the case when s_t is one of the explanatory variables and only the first element of the vector θ is different from 0. This means that the extreme regimes of model (2.4) differ only in the constant term. Obviously, the test has no power if $\theta = (\theta_0, 0, \ldots, 0)'$, $\theta_0 \neq 0$, as η_1 is equal to zero even under the alternative hypothesis of a nonlinear model. The problem can be solved by using a higher-order Taylor approximation of the transition function. Since the second partial derivative of the transition function G_1^* is equal to zero at $\gamma = 0$, it seems sensible to use the third-order Taylor polynomial,

$$T_3(\gamma, c; s_t) = \delta_0 + \delta_1\, s_t + \delta_2\, s_t^2 + \delta_3\, s_t^3 + R_3(\gamma, c; s_t). \tag{3.12}$$

After replacing the transition function G with T_3 in equation (2.4) and rearranging the terms, one obtains

$$y_t = x_t'\eta_0 + (\tilde{x}_t' s_t)\eta_1 + (\tilde{x}_t' s_t^2)\eta_2 + (\tilde{x}_t' s_t^3)\eta_3 + u_t^*, \tag{3.13}$$

with $u_t^* = x_t'\theta R_3(\gamma, c; s_t) + u_t$. The null hypothesis of linearity,

$$H_0''' : \eta_1 = \eta_2 = \eta_3 = 0, \tag{3.14}$$

is tested as for equations (3.7) and (3.11), with the LM test of the form (3.9). Only the vector w_t has to be changed to

$$w_t = (\tilde{x}_t' s_t, \tilde{x}_t' s_t^2, \tilde{x}_t' s_t^3)'.$$

The computed test statistic is asymptotically χ^2-distributed with 3p degrees of freedom. The alternative hypothesis to (3.14) is obviously H_1''' : at least one of the η_i, $i = 1, 2, 3$, is not equal to 0.

So far, we have only discussed linearity testing for the LSTR1 model under the alternative hypothesis. In case of the ESTR model, the first-order Taylor approximation of the transition function G_3^* leads to a polynomial of degree 2 in the transition variable s_t. Escribano and Jordá [20] argue that the first-order Taylor expansion is not sufficient to reflect the characteristics of the original transition function. They propose using the second-order Taylor polynomial, instead. Consequently, the degree of freedom of the resulting LM test statistic amounts to 2p in the first case and 4p in the second case.

The linearity test for the LSTR2 model is developed in a similar way. One could use the Taylor approximation of order one, which yields a polynomial of degree 2 in the transition function s_t and the LM test statistic with 2p degrees of freedom, or the Taylor approximation of order three, which leads to a polynomial of degree 6 and the test statistic with 6p degrees of freedom. The second partial derivative of the transition function G_2^* with respect to γ is equal to 0 at $\gamma = 0$, therefore we have left out the Taylor expansion of order 2. Because of the high number of coefficients that have to be tested in the second case, it is better to use only the first-order Taylor approximation at the expansion point $\gamma = 0$.

We shall restrict our attention to first-order Taylor polynomials for ESTR and LSTR2 transition functions, to obtain a general linearity test that covers all 3 STR models. Namely, we could run auxiliary regression (3.13) and test the null hypothesis H_0''' : $\eta_1 = \eta_2 = \eta_3 = 0$ in all 3 cases. ESTR and LSTR2 transition functions would lead to regression (3.13) with $\eta_3 = 0$, but the proposed linearity test would still have power against ESTR and LSTR2 models under the alternative.

If the number of restrictions under the null hypothesis is large when compared to the sample size, the asymptotic χ^2 distribution is likely to be a poor approximation for the actual small-sample distribution of the LM test statistic. In this case, an F - approximation of the test statistic works much better, as the empirical size of the test remains close to the nominal size while the power is good. For this reason, Granger and Teräsvirta [27] suggest using the F - test for testing linearity against STR. The description of the F - approximation of the LM test in given in section 3.6.5. The test is performed in 3 steps:

1. After regressing y_t on x_t, use the obtained residuals to define the residual sum of squares $SSR_0 = \sum_{t=1}^{T} \hat{u}_t^2$.

2. Compute the residual sum of squares (SSR_1) from regressing $\hat{u}_t$ on $(x_t', \tilde{x}_t' s_t, \tilde{x}_t' s_t^2, \tilde{x}_t' s_t^3)'$.

3. Calculate the F statistic as

$$F = \frac{(SSR_0 - SSR_1)/(3p)}{SSR_1/(T - 4p - 1)}. \tag{3.15}$$

   Under null hypothesis (3.14), the test statistic is F-distributed with 3p and T-4p-1 degrees of freedom.

The above theory covers the possibility when s_t is one of the explanatory variables, i.e. s_t is an element of x_t. If this is not the case, the auxiliary regression to be run is

$$y_t = x_t'\eta_0 + (x_t' s_t)\eta_1 + (x_t' s_t^2)\eta_2 + (x_t' s_t^3)\eta_3 + u_t^*, \tag{3.16}$$

The resulting F statistic has 3(p+1) and T-4p-4 degrees of freedom.

3.2 Model specification

3.2.1 Choosing the transition variable

When developing the linearity test, we have always assumed that the transition variable s_t is known. But the choice of the transition variable is not straightforward, since the underlying economic theory often gives no clues as to which variable should be taken for the transition variable under the alternative. If we are choosing s_t only from the the set of explanatory variables, then we can specify the transition variable as a linear combination of the elements of $\tilde{x}_t$ of the form

$$s_t = a'\tilde{x}_t = (0, \ldots, 0, 1, 0, \ldots, 0)'\tilde{x}_t. \tag{3.17}$$

The position of the only nonzero element of the vector a determines the transition variable, but it is not known when performing the linearity test. The procedure is explained in Luukkonen, Saikkonen and Teräsvirta [50] and involves approximating the transition function with its third order Taylor polynomial, substituting it for G in equation (2.4) and replacing s_t with $a'\tilde{x}_t$. This yields the auxiliary regression

$$y_t = x_t'\eta_0 + \sum_{i=1}^{p}\sum_{j=i}^{p}\eta_{1ij}x_{ti}x_{tj} + \sum_{i=1}^{p}\sum_{j=1}^{p}\eta_{2ij}x_{ti}x_{tj}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p}\eta_{3ij}x_{ti}x_{tj}^3 + u_t^*, \quad (3.18)$$

with the null hypothesis of linearity of the form

$$\begin{aligned} H_0 \quad : \quad & \eta_{1ij} = 0, \quad i = 1, 2, \ldots, p, \ j = i, i+1, \ldots, p, \\ & \eta_{2ij} = \eta_{3ij} = 0, \quad i = 1, 2, \ldots, p, \ j = 1, 2, \ldots, p. \end{aligned} \quad (3.19)$$

The resulting LM statistic is asymptotically χ^2-distributed with degree of freedom equal to $\frac{p(p+1)}{2} + 2p^2$. As this expression increases fast with the growing p, a restricted version of (3.18) is usually used in practice:

$$y_t = x_t'\eta_0 + \sum_{i=1}^{p}\sum_{j=i}^{p}\eta_{1ij}x_{ti}x_{tj} + \sum_{j=1}^{p}\eta_{3j}x_{tj}^3 + u_t^*. \quad (3.20)$$

For the null hypothesis

$$\begin{aligned} H_0 \quad : \quad & \eta_{1ij} = 0, \quad i = 1, 2, \ldots, p, \ j = i, i+1, \ldots, p, \\ & \eta_{3j} = 0, \quad j = 1, 2, \ldots, p, \end{aligned} \quad (3.21)$$

the degree of freedom is reduced to the value of $\frac{p(p+1)}{2} + p$. This is a general test of linearity against the alternative of a STR model, where the transition variable does not have to be known a priori. If the choice of the transition variable is restricted to a subset of the elements of $\tilde{x}_t$, the procedure can be modified accordingly.

An alternative way of proceeding is to test the null hypothesis of linearity with auxiliary regression (3.13) for each of the possible transition variables in turn. The candidates for the transition variable are usually the explanatory variables and the time trend. If the null is rejected for more than one variable, the variable with the strongest rejection of linearity (i.e. with the lowest p-value) is chosen for the transition variable. This intuitive and heuristic procedure can be justified by observing that the test is most powerful when the alternative hypothesis is correctly specified, and this is achieved for the "right" transition variable. It has to be emphasized that one cannot control the overall significance level of the linearity test for this heuristic procedure, since several individual tests have to be performed. If the overall significance level is important, one should test the null of linearity with a test based on auxiliary regression (3.18) or (3.20), where the transition variable does not have to be determined a priori.

3.2.2 Choosing the transition function

If the transition variable has already been decided upon, the next step in the modeling process consists of choosing the transition function. For each of the 3 types of the transition function in model (2.4), Granger and Teräsvirta [27] expressed the vectors η_1, η_2 and η_3 of auxiliary regression (3.13), namely

$$y_t = x_t'\eta_0 + (\tilde{x}_t' s_t)\eta_1 + (\tilde{x}_t' s_t^2)\eta_2 + (\tilde{x}_t' s_t^3)\eta_3 + u_t^*, \tag{3.22}$$

as functions of the parameters γ, c (or parameters c_1 and c_2 for LSTR2 model), $\theta = (\theta_0, \theta_1, \ldots, \theta_p)' = (\theta_0, \tilde{\theta}')'$, and the values of the first three partial derivatives of the transition function G at the point $\gamma = 0$. They concluded the following:

(i) $\eta_3 = 0$ for ESTR and LSTR2 models assuming that the first-order Taylor approximation of the transition function was used. For LSTR1 model, η_3 is generally not equal to 0 (unless $\tilde{\theta} = 0$).

(ii) $\eta_2 = 0$ in case of an LSTR1 model with $\theta_0 = c = 0$. For ESTR and LSTR2 models, η_2 is usually different from 0.

(iii) $\eta_1 = 0$ for ESTR models with $\theta_0 = c = 0$ and for LSTR2 models with $\theta_0 = c_1 = c_2 = 0$. In case of an LSTR1 model, η_1 is generally not equal to 0.

Consequently, Granger and Teräsvirta [27] based their decision rule on a sequence of nested hypotheses:

$$\begin{aligned} H_{04} &: \eta_3 = 0, \\ H_{03} &: \eta_2 = 0 |\, \eta_3 = 0, \\ H_{02} &: \eta_1 = 0 |\, \eta_2 = \eta_3 = 0. \end{aligned} \tag{3.23}$$

The 3 hypotheses are tested with a sequence of F-tests named F4, F3 and F2, respectively. If the rejection of the hypothesis H_{03} is the strongest, Granger and Teräsvirta advise choosing the LSTR2 or the ESTR model. In the practice, one usually chooses the LSTR2 model and additionally tests the hypothesis $c_1 = c_2$ after estimation. If it cannot be rejected, it seems better to choose the LSTR2 model, otherwise ESTR should be chosen. In case of the strongest rejection of the hypothesis H_{04} or H_{02}, LSTR1 is selected as the appropriate model.

There is also an alternative to the previously described decision rule, which requires a high computational load. For fixed values of γ and c, model (2.4) is linear in parameters and can be estimated by ordinary least squares. In the first step, the parameters γ and c run through a two-dimensional grid for LSTR1 models, or the parameters γ, c_1 and c_2 take the values from a three-dimensional grid in case of LSTR2 models. At each point of the grid the obtained linear model (linear in parameters) is estimated. The parameters c, c_1 and c_2 are allowed to take only the values between the observed minimum and the observed maximum of the transition variable s_t. In the second step, one chooses the estimated model with the best fit from the set of all LSTR1 and LSTR2 models. As the measure of the best fit, standard error of regression can be used. If the LSTR2 model was chosen, the hypothesis $c_1 = c_2$ is tested and if it is not rejected, ESTR model is selected. The same grid-search procedure is used for setting the starting values for the estimation of STR models.

Teräsvirta [73] conducted a series of simulation experiments to investigate the properties of the proposed heuristic specification strategy for choosing the transition variable and the transition function. The study was conducted for smooth transition autoregressive (STAR) models in the univariate setting. Different types of STAR models were examined and their parameters were varied. The "true" transition variable was the lagged endogenous variable y_{t-d}, where the delay parameter d ran from 1 to 5. For each d the linearity test was performed for every possible transition variable in turn (i.e. for $y_{t-1}, y_{t-2}, \ldots, y_{t-5}$) and the variable with the lowest p-value was chosen. The empirical size of the overall linearity test was 3 to 4 % when the nominal size was 5 %. The results of the simulation study justified the heuristic specification procedure and also showed that the power of the linearity test is better for higher γ values and for lower values of the delay parameter d. The decision rule for choosing the type of the transition function was tested for distinguishing between LSTAR1 and ESTAR models. It works best when the number of observations of the transition variable that lie below c is about the same as the number of observations above c. The performance of the rule improves with the sample size.

3.3 Estimation of STR models

The specified STR model is usually estimated with nonlinear least squares or with maximum likelihood estimation under the assumption of normally distributed errors. Both methods are equivalent in this case. As already described in the previous section, the grid search can be applied to obtain initial estimates of the parameters. Nonlinear optimization procedures are used to maximize the log-likelihood or to minimize the sum of squared residuals. The applied STR models from the next chapters are estimated with the Gauss package or with EViews. Several nonlinear optimization algorithms are available in Gauss. For example, the Newton - Raphson algorithm, the Broyden - Fletcher - Goldfarb - Shanno (BFGS) algorithm, the steepest descent algorithm and the Davidon - Fletcher - Powell (DFP) algorithm are all implemented in the Gauss library Optmum. A complete exposition on nonlinear optimization is given in Fletcher [22] and in Nocedal and Wright [57].

An additional remark should be made on the slope parameter γ of the transition function. The magnitude of the parameter γ depends on the magnitude of the transition variable s_t and is therefore not scale-free. The numerical optimization is more stable if the exponent of the transition function is standardized prior to optimization. In other words, it is advisable to divide γ by the sample standard deviation (in case of LSTR1 models) or by the sample variance (for ESTR and LSTR2 models) of the transition variable. In this way the magnitude of the slope parameter is brought closer to the magnitude of other parameters.

3.4 Misspecification tests

After estimating the specified STR model, one has to perform the specification tests to check if the underlying assumptions hold. The tests described in this section were first developed by Eitrheim and Teräsvirta [18] in a univariate setting, i.e. for smooth transition autoregressive (STAR) models. The generalization to STR models is straightforward, provided that:

1. the parameters of the smooth transition models are estimated consistently,

2. the estimates are asymptotically normally distributed.

Extensive discussion on both conditions can be found in Wooldridge [83] and in Escribano and Mira [21]. The description of all misspecification tests for STR models is given in Teräsvirta [72].

3.4.1 Error autocorrelation

The usual tests of Ljung and Box cannot be applied in this case, since their asymptotic distribution under the null hypothesis is unknown when they are calculated from estimated residuals of a STAR or STR model.

Let us start the development of the test by considering a general STR model with the functional form

$$\begin{aligned} y_t &= M(x_t; \psi) + u_t, \\ M(x_t; \psi) &= x_t' \varphi + (x_t' \theta) G_i(\gamma, c; s_t), \quad i = 1, 2, 3, \end{aligned} \tag{3.24}$$

and AR(q) errors

$$u_t = a_1 u_{t-1} + a_2 u_{t-2} + \cdots + a_q u_{t-q} + \varepsilon_t = a' v_t + \varepsilon_t, \tag{3.25}$$

where $a = (a_1, a_2, \cdots, a_q)'$, $v_t = (u_{t-1}, u_{t-2}, \cdots, u_{t-q})'$, $\psi = (\varphi', \theta', \gamma, c)'$, and the errors ε_t are independently identically distributed with $\varepsilon_t \sim N(0, \sigma^2)$. Using the lag operator L, one can rewrite equation (3.25) as

$$\varepsilon_t = \left(1 - \sum_{i=1}^{q} a_i L^i\right) u_t =: p(L) u_t. \tag{3.26}$$

The null hypothesis of no error autocorrelation up to lag order q is therefore $H_0 : a = 0$ and the alternative hypothesis is $H_1 : a \neq 0$. The roots of the polynomial p(z) are assumed to lie outside the unit circle.

The derivation of the test is given in a more general setting, when the function M is only supposed to be at least twice continuously differentiable with respect to its parameters. Eitrheim and Teräsvirta [18] impose an additional assumption that under the null hypothesis $\{y_t\}$ is stationary and ergodic. This assumption is important for consistent estimation of parameters by nonlinear least squares (NLS). A comprehensive and technical discussion on the necessary and sufficient conditions for consistent NLS estimation can be found in Klimko and Nelson [44].

Assume that the starting values $y_0, y_{-1}, \dots, y_{-q+1}$ and $x_0, x_{-1}, \dots, x_{-q+1}$ are known. Since

$$\varepsilon_t = u_t - a' v_t = \Big(y_t - M(x_t;\psi)\Big) - \sum_{j=1}^{q} a_j \Big(y_{t-j} - M(x_{t-j};\psi)\Big), \qquad (3.27)$$

the conditional log - likelihood function is

$$\begin{aligned} l &= c - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2 & (3.28)\\ &= c - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\cdot \\ &\cdot \sum_{t=1}^{T}\Big(y_t - M(x_t;\psi) - \sum_{j=1}^{q} a_j y_{t-j} + \sum_{j=1}^{q} a_j M(x_{t-j};\psi)\Big)^2. \end{aligned}$$

The parameters of the model are stacked in the vector $\eta = (\psi', a')'$, with the exception of σ^2. As the information matrix is block diagonal, we can consider σ^2 fixed when computing the test statistic. A similar argument is explained in the derivation of the LM test in section 3.6.2. The first partial derivatives of the log - likelihood function are

$$\begin{aligned} \frac{\partial l}{\partial a_j} &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\varepsilon_t\Big(y_{t-j} - M(x_{t-j};\psi)\Big), \quad j = 1,2,\dots,q, \\ \frac{\partial l}{\partial \psi} &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\varepsilon_t\Big(\frac{\partial M(x_t;\psi)}{\partial\psi} - \sum_{j=1}^{q} a_j \frac{\partial M(x_{t-j};\psi)}{\partial\psi}\Big). & (3.29) \end{aligned}$$

Taking into account that for STR models

$$\begin{aligned} \frac{\partial M(x_t;\psi)}{\partial\psi} &= \Big(\frac{\partial M}{\partial\varphi'}, \frac{\partial M}{\partial\theta'}, \frac{\partial M}{\partial\gamma}, \frac{\partial M}{\partial c}\Big)' \\ &= \Big(x_t', x_t' G_i, \frac{\partial G_i}{\partial\gamma}\,\theta' x_t, \frac{\partial G_i}{\partial c}\,\theta' x_t\Big)', \quad i = 1,2,3, & (3.30) \end{aligned}$$

the necessary partial derivatives can be computed for the 3 types of the transition function:

1. LSTR1

$$
\begin{aligned}
G_1(\gamma, c; s_t) &= \frac{1}{1+e^{-\gamma(s_t-c)}} \qquad \gamma > 0, \\
\frac{\partial G_1(\gamma, c; s_t)}{\partial \gamma} &= \frac{e^{-\gamma(s_t-c)}(s_t-c)}{\left(1+e^{-\gamma(s_t-c)}\right)^2} \\
&= \frac{(s_t-c)}{\left(e^{\frac{\gamma}{2}(s_t-c)}+e^{-\frac{\gamma}{2}(s_t-c)}\right)^2} \\
\frac{\partial G_1(\gamma, c; s_t)}{\partial c} &= -\frac{\gamma}{\left(e^{\frac{\gamma}{2}(s_t-c)}+e^{-\frac{\gamma}{2}(s_t-c)}\right)^2}
\end{aligned} \tag{3.31}
$$

2. LSTR2

$$
\begin{aligned}
G_2(\gamma, c_1, c_2; s_t) &= \frac{1}{1+e^{-\gamma(s_t-c_1)(s_t-c_2)}} \qquad \gamma > 0, \quad c_1 \leq c_2 \\
\frac{\partial G_2(\gamma, c_1, c_2; s_t)}{\partial \gamma} &= \frac{(s_t-c_1)(s_t-c_2)}{\left(e^{\frac{\gamma}{2}(s_t-c_1)(s_t-c_2)}+e^{-\frac{\gamma}{2}(s_t-c_1)(s_t-c_2)}\right)^2} \\
\frac{\partial G_2(\gamma, c_1, c_2; s_t)}{\partial c_1} &= \frac{-\gamma(s_t-c_2)}{\left(e^{\frac{\gamma}{2}(s_t-c_1)(s_t-c_2)}+e^{-\frac{\gamma}{2}(s_t-c_1)(s_t-c_2)}\right)^2} \\
\frac{\partial G_2(\gamma, c_1, c_2; s_t)}{\partial c_2} &= \frac{-\gamma(s_t-c_1)}{\left(e^{\frac{\gamma}{2}(s_t-c_1)(s_t-c_2)}+e^{-\frac{\gamma}{2}(s_t-c_1)(s_t-c_2)}\right)^2}
\end{aligned} \tag{3.32}
$$

3. ESTR

$$
\begin{aligned}
G_3(\gamma, c; s_t) &= 1-e^{-\gamma(s_t-c)^2} \qquad \gamma > 0 \\
\frac{\partial G_3(\gamma, c; s_t)}{\partial \gamma} &= (s_t-c)^2 e^{-\gamma(s_t-c)^2} \\
\frac{\partial G_3(\gamma, c; s_t)}{\partial c} &= -2\gamma(s_t-c)e^{-\gamma(s_t-c)^2}
\end{aligned} \tag{3.33}
$$

Let us denote $\frac{\partial M(x_t;\hat{\psi})}{\partial \hat{\psi}}$ by $\hat{z}_t$. Using formula (3.101) for the computation of the Lagrange multiplier test statistic yields

$$\begin{aligned} LM &= \frac{1}{\hat{\sigma}^2}\left(\sum_{t=1}^{T}\hat{u}_t\hat{v}_t\right)'\hat{I}^{-1}\left(\sum_{t=1}^{T}\hat{u}_t\hat{v}_t\right), \\ \hat{I}^{-1} &= \left(\sum_{t=1}^{T}\hat{v}_t\hat{v}_t' - \left(\sum_{t=1}^{T}\hat{v}_t\hat{z}_t'\right)\left(\sum_{t=1}^{T}\hat{z}_t\hat{z}_t'\right)^{-1}\left(\sum_{t=1}^{T}\hat{z}_t\hat{v}_t'\right)\right)^{-1}. \end{aligned} \tag{3.34}$$

We have already observed that under H_0 consistent estimators of (3.29) are given by

$$\begin{aligned} \frac{\partial \hat{l}}{\partial a} &= \frac{1}{\hat{\sigma}^2}\sum_{t=1}^{T}\hat{u}_t\hat{v}_t, \\ \frac{\partial \hat{l}}{\partial \psi} &= \frac{1}{\hat{\sigma}^2}\sum_{t=1}^{T}\hat{u}_t\hat{z}_t, \end{aligned} \tag{3.35}$$

where $\hat{u}_{t-j} = y_{t-j} - M(x_{t-j};\hat{\psi})$, $j = 0, 1, \ldots, q$, $\hat{v}_t = (\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-q})'$, $\hat{\psi}$ is the NLS estimator of ψ under H_0 and $\hat{\sigma} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2$. Obviously, LM is asymptotically χ^2 - distributed with q degrees of freedom.

Eitrheim and Teräsvirta [18] suggest using the F - version of the test statistic, because its empirical size is close to the nominal size and its power is good. They investigated the size and power properties by means of a simulation study. The proposed F - test (discussed in section 3.6.5) is usually carried out in 3 steps:

1. After estimating the model by NLS (under the assumption of no error autocorrelation), use the obtained residuals to define the residual sum of squares $SSR_0 = \sum_{t=1}^{T}\hat{u}_t^2$.

2. Compute the residual sum of squares (SSR_1) from the auxiliary regression of $\hat{u}_t$ on $\hat{v}_t$ and $\hat{z}_t$.

3. If the dimension of the vector $\hat{z}_t$ is denoted by n, then the test statistic is given by
$$F = \frac{(SSR_0 - SSR_1)/q}{SSR_1/(T-n-q)}. \tag{3.36}$$

3.4.2 Testing for remaining nonlinearity

After a smooth transition regression model has been estimated, a natural question arises: Does the specified model adequately capture the nonlinear features of the observed time series? Intuitively, one would probably come up with the idea of introducing an additional additive nonlinear term of the smooth transition kind into model (2.4):

$$y_t = x_t'\varphi + (x_t'\theta) \cdot G(\gamma_1, c_1; s_t) + (x_t'\psi) \cdot H(\gamma_2, c_2; r_t) + u_t. \tag{3.37}$$

To develop the test of no remaining nonlinearity, one may obviously perform the same steps as for the original linearity test. The null hypothesis $\gamma_2 = 0$ suffers from the identification problem, which can be solved as before by approximating the transition function H with its Taylor polynomial of order 3 around $\gamma_2 = 0$. After rearranging some of the terms in equation (3.37), the model translates to

$$y_t = x_t'\beta_0 + (x_t'\theta) \cdot G(\gamma_1, c_1; s_t) + (\tilde{x}_t' r_t)\beta_1 + (\tilde{x}_t' r_t^2)\beta_2 + (\tilde{x}_t' r_t^3)\beta_3 + u_t^*. \tag{3.38}$$

The null hypothesis of no remaining nonlinearity is $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. To compute the test statistic from auxiliary regression (3.38), the 3 steps of the no error autocorrelation test are performed with only one modification. Namely, in the second step $\hat{v}_t$ now stands for $(\tilde{x}_t' r_t, \tilde{x}_t' r_t^2, \tilde{x}_t' r_t^3)'$. The resulting F-statistic has 3p and T-4p-1 degrees of freedom.

The vector $\tilde{x}_t$ does not have to include all of the components of x_t (with the exception of the constant term), but only a suitable subset. In other words, some of the parameters may be equal to zero.

As noted by Teräsvita [72], it is not sensible to carry out the test of no remaining nonlinearity for several possible transition variables in turn. One should better use the generalized version of the test, derived under the assumption that the transition variable r_t is a linear combination of the components of $\tilde{x}_t$. In this case, the transition variable is of the form $a'\tilde{x}_t$, where the vector a contains only one nonzero element. This idea has already been described and used in the section about the linearity test.

The test of no remaining nonlinearity can be modified so as to allow a more general additive nonlinear term under the alternative hypothesis:

$$y_t = x_t'\varphi + (x_t'\theta) \cdot G(\gamma, c; s_t) + K(x_t) + u_t. \tag{3.39}$$

The third order Taylor expansion of the function $K(x_t)$ exists when K is at least three times continuously differentiable. If so, we write down the expansion around the point $x_t = x_t^0$:

$$K(x_t) = \kappa_0' x_t + \sum_{i=1}^{p}\sum_{j=i}^{p} \kappa_{ij} x_{ti} x_{tj} + \sum_{i=1}^{p}\sum_{j=i}^{p}\sum_{l=j}^{p} \kappa_{ijl} x_{ti} x_{tj} x_{tl} + R_K(x_t). \quad (3.40)$$

After substituting $K(x_t)$ with (3.40) in (3.39), one obtains

$$\begin{aligned} y_t &= x_t'\beta_0 + (x_t'\theta) \cdot G(\gamma, c; s_t) + \\ &+ \sum_{i=1}^{p}\sum_{j=i}^{p} \kappa_{ij} x_{ti} x_{tj} + \sum_{i=1}^{p}\sum_{j=i}^{p}\sum_{l=j}^{p} \kappa_{ijl} x_{ti} x_{tj} x_{tl} + u_t^* \end{aligned} \quad (3.41)$$

Under the null hypothesis of no remaining nonlinearity the function $K(x_t)$ is linear, hence

$$\begin{aligned} H_0 \ :& \ \kappa_{ij} = 0,\ i = 1, \ldots, p,\ j = i, \ldots, p; \\ & \ \kappa_{ijl} = 0,\ i = 1, \ldots, p,\ j = i, \ldots, p,\ l = j, \ldots, p. \end{aligned} \quad (3.42)$$

The hypothesis can be tested by an F-test. One has to be careful not to ignore the fact that many parameters are restricted under the null hypothesis. Consequently, the test should not be used if p is large and the sample is small, because of low power.

3.4.3 Parameter constancy test

As parameter constancy is assumed when the parameters of a smooth transition model are tested, this assumption should be verified in order to avoid misspecification. In linear models, most of the parameter tests are performed to test for a single structural break (or a finite number of breaks) in the sample. Lin and Teräsvirta [48], on the other hand, consider deterministic change in parameters over time as an alternative to the hypothesis of parameter constancy. In their paper, the alternative hypothesis is a parametric hypothesis with smoothly changing parameters. Another kind of parameter constancy tests can also be found in the literature, where the parameters under the alternative are stochastic. For example, Nyblom and Mäkeläinen [59], Nyblom [58] and Hansen [31] are concerned with random walk - type parameters under the alternative. Following Teräsvirta [72], this subsection generalizes the results obtained by Lin and Teräsvirta [48] for STAR models.

We start with the STR model

$$y_t = (x_t^0)'\varphi^0(t) + (x_t^1)'\theta^0(t)G(\gamma, c; s_t) + u_t, \tag{3.43}$$

where the vector x_t^0 of length p_0 contains only those elements of the vector x_t whose coefficients are not assumed zero a priori, and the vector x_t^1 of length p_1 in the nonlinear part is defined in a similar way. Suppose that the parameters of the transition function (γ and c) do not change over time, whereas the parameter vectors φ^0 and θ^0 are time - dependent. More specifically, assume that the time - dependency is described by the functions

$$\begin{aligned} \varphi^0(t) &= \varphi^0 + \lambda_1 H(\gamma_1, c_1; t), \\ \theta^0(t) &= \theta^0 + \lambda_2 H(\gamma_1, c_1; t), \end{aligned} \tag{3.44}$$

with φ^0 and λ_1 of the same dimension as the vector x_t^0, while θ^0 and λ_2 are of the same dimension as the vector x_t^1. The parameter vector c_1 contains 1, 2 or 3 parameters. For the function H, 3 possibilities were proposed by Jansen and Teräsvirta [38]:

$$\begin{aligned} H_1(\gamma_1, c_1; t) &= \frac{1}{1 + e^{-\gamma_1(t-c_1)}} - \frac{1}{2}, \\ H_2(\gamma_1, c_1; t) &= \frac{1}{1 + e^{-\gamma_1(t-c_{11})(t-c_{12})}} - \frac{1}{2}, \\ H_3(\gamma_1, c_1; t) &= \frac{1}{1 + e^{-\gamma_1(t-c_{11})(t-c_{12})(t-c_{13})}} - \frac{1}{2}, \end{aligned} \tag{3.45}$$

under the usual assumptions of $\gamma > 0$ and $c_{11} \leq c_{12} \leq c_{13}$. The parameter γ is the slope parameter indicating how fast the parameters of the model change.

The function H_1 enables a single structural break, for $\gamma_1 \to \infty$. By choosing H_2 with $c_{11} < c_{12}$ two structural breaks can be modeled, whereas H_3 is used in case of asymmetrically and nonmonotonically changing parameters.

When discussing the parameter constancy test, the transition function is set to H_3, since the functions H_1 and H_2 can be considered as special cases of H_3. Our goal is to test the null hypothesis of parameter constancy

$$H_0 : \gamma_1 = 0 \tag{3.46}$$

against the alternative hypothesis $H_1 : \gamma_1 > 0$. Again, the problem of nuisance parameters is encountered under H_0, and is also dealt with in the same way as for the linearity tests. In the first step, H_3 is approximated by its Taylor polynomial of order one around $\gamma = 0$, which gives

$$\begin{aligned} H_3(\gamma_1, c_1; t) &= \delta_0 + \delta_1(t - c_{11})(t - c_{12})(t - c_{13}) + R_3(\gamma_1, c_1; t) \\ &= d_0 + d_1 t + d_2 t^2 + d_3 t^3 + R_3(\gamma_1, c_1; t) \\ &= T_3(\gamma_1, c_1; t) + R_3(\gamma_1, c_1; t). \end{aligned} \tag{3.47}$$

After approximating $\varphi^0(t)$ and $\theta^0(t)$ in model (3.43) with the help of (3.47), one obtains

$$\begin{aligned} y_t &= (x_t^0)'\beta_0 + (x_t^0 t)'\beta_1 + (x_t^0 t^2)'\beta_2 + (x_t^0 t^3)'\beta_3 + \\ &+ \Big((x_t^1)'\beta_4 + (x_t^1 t)'\beta_5 + (x_t^1 t^2)'\beta_6 + (x_t^1 t^3)'\beta_7\Big) G(\gamma, c; s_t) + u_t^*, \end{aligned} \tag{3.48}$$

where $u_t^* = u_t + \Big((x_t^0)'\lambda_1 + (x_t^1)'\lambda_2 G(\gamma, c; s_t)\Big) R_3(\gamma_1, c_1; t)$ and β_j is equal to γb_j, for $j = 1, 2, 3, 5, 6, 7$. Obviously, the parameters are constant when the hypothesis

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_5 = \beta_6 = \beta_7 = 0 \tag{3.49}$$

holds. To derive the asymptotic normality of the estimated parameter vector and consequently the asymptotic χ^2-distribution of the LM test, it is important to notice that the partial derivatives $\frac{\partial G}{\partial \gamma} = g_\gamma(t)$ and $\frac{\partial G}{\partial c} = g_c(t)$ are bounded functions. These partial derivatives were computed in subsection 3.4.1 and are given in equations (3.31), (3.32) and (3.33) for the transition functions G_1, G_2 and G_3, respectively. Eitrheim and Teräsvirta [18] established the asymptotic theory using the fact that the previously mentioned functions $g_\gamma(t)$ and $g_c(t)$ are bounded (see [18] for details). The LM test statistic is again computed from equation (3.101), namely

$$LM = \frac{1}{\hat{\sigma}^2} \left(\sum_{t=1}^{T} \hat{w}_t \hat{u}_t \right)' \left(\hat{M}_{22} - \hat{M}_{21} \hat{M}_{11}^{-1} \hat{M}_{12} \right)^{-1} \left(\sum_{t=1}^{T} \hat{w}_t \hat{u}_t \right), \tag{3.50}$$

with

$$\begin{aligned} \hat{z}_t &= \Big((x_t^0)', \big(x_t^1 G(\hat{\gamma}, \hat{c}; s_t)\big)', \hat{g}_\gamma(t), \hat{g}_c(t)\Big)', && (3.51) \\ \hat{w}_t &= \Big((x_t^0 t)', (x_t^0 t^2)', (x_t^0 t^3)', \big(x_t^1 t \hat{G}\big)', \big(x_t^1 t^2 \hat{G}\big)', \big(x_t^1 t^3 \hat{G}\big)'\Big)', && (3.52) \\ \hat{G} &= G(\hat{\gamma}, \hat{c}; s_t), \end{aligned}$$

$\hat{M}_{11} = \sum_{t=1}^{T} \hat{z}_t \hat{z}_t'$, $\hat{M}_{22} = \sum_{t=1}^{T} \hat{w}_t \hat{w}_t'$, $\hat{M}_{12} = \hat{M}_{21}' = \sum_{t=1}^{T} \hat{z}_t \hat{w}_t'$, $\hat{u}_t$ is the residual estimated under the null hypothesis and $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2$. The test statistic is asymptotically χ^2-distributed with $3(p_0 + p_1)$ degrees of freedom. Again, using the F-test instead of the LM-test is recommended. The degrees of freedom of the F statistic are $3(p_0 + p_1)$ and $T - 4(p_0 + p_1)$, respectively.

For the function H_2, the null hypothesis is

$$H_0 : \beta_1 = \beta_2 = \beta_5 = \beta_6 = 0, \tag{3.53}$$

and for H_3,

$$H_0 : \beta_1 = \beta_5 = 0. \tag{3.54}$$

Sometimes one may wish to test the parameter constancy only for a subset of the parameters of a given model. For this purpose, the vectors of regressors x_t^0 and x_t^1 are partitioned into vectors x_t^{01}, x_t^{02} and x_t^{11}, x_t^{12}, respectively. Partitioning the parameter vectors φ and θ in (3.43) accordingly leads to

$$y_t = (x_t^{01})'\varphi_1 + (x_t^{02})'\varphi_2^0(t) + \Big((x_t^{11})'\theta_1 + (x_t^{12})'\theta_2^0(t)\Big)G(\gamma, c; s_t) + u_t. \tag{3.55}$$

The parameter constancy test can now be carried out only for the coefficients of the vectors x_t^{02} and x_t^{12}.

Lin and Teräsvirta [48] suggest slightly different functions to characterize parameter change. Namely, the polynomials in the exponent of H_1 and H_3 are replaced by general polynomials of order 1 and 3, respectively, whereas the function replacing H_2 is not a logistic function and its exponent is a square of a linear polynomial. In the notation of Lin and Teräsvirta [48],

$$\begin{aligned} F_k(\gamma_1, \alpha; t) &= \frac{1}{1 + e^{-\gamma_1 (t^k + \alpha_1 t^{k-1} + \ldots + \alpha_{k-1} t + \alpha_k)}}, \qquad k = 1, 3, \\ F_k(\gamma_1, \alpha; t) &= 1 - e^{-\gamma_1 (t-\alpha)^2}, \qquad k = 2. \end{aligned} \tag{3.56}$$

Replacing the function F_3 by its first order Taylor approximation in (3.43) yields a reparametrized model

$$y_t = s_t'\lambda + (s_t \otimes \tilde{x}_t)'\delta + u_t, \tag{3.57}$$

where $s_t = (1, t, t^2, t^3)'$, $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)'$, $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-m}, z_{t1}, \ldots, z_{tn})'$, $(s_t \otimes \tilde{x}_t) = (\tilde{x}_t', t\tilde{x}_t', t^2\tilde{x}_t', t^3\tilde{x}_t')'$ and $\delta = (\delta_0', \delta_1', \delta_2', \delta_3')'$, while the vectors δ_i, $i = 0, 1, 2, 3$ are of length $m + n$. Instead of $\gamma_1 = 0$, the null hypothesis of parameter constancy can be written as

$$H_0 : \lambda_1 = \lambda_2 = \lambda_3 = \delta_1 = \delta_2 = \delta_3 = 0. \tag{3.58}$$

The usual LM test statistic for testing linear restrictions can easily be derived and used to test hypothesis (3.58). But Lin and Teräsvirta [48] recommend using the F-statistic instead. They argue that the F-statistic has better small sample properties, since its empirical size is close to the nominal size, whereas for the χ^2-test the difference between the empirical size and the nominal size can be substantial in small samples. The tests are called LM1, LM2 and LM3 (for $k = 1, 2, 3$, respectively), although they are not Lagrange multiplier tests.

The small sample properties of the 3 tests are studied by means of Monte Carlo experiments and the results are compared to those of the CUSUM test, CUSUMQ test and the test with the random walk alternative hypothesis of Nyblom [58], called the *N-test*. In case of a single structural break in the generated data, the highest power is achieved by LM tests, but the power of the N-test is also satisfactory. The results of the CUSUM and CUSUMQ test are not encouraging. While the CUSUM test preforms worse if the change of the parameters occurs late in the sample, the CUSUMQ test demonstrates low power in general. A case of a double breakpoint is also investigated. After the second breakpoint, the data generating process is the same as before the first breakpoint. Thus, the change in the parameters is not monotonic and the power properties of LM1 are inferior to those of LM2 and LM3, as expected. The N-test behaves similarly as the LM1 test and the CUSUM test performs even worse, signaling that it is difficult to detect a double break with the help of CUSUM.

If one wishes to estimate the parameters under the alternative hypothesis, a suitable value of k has to be chosen first. Unfortunately, the economic theory does not provide any guidance in the selection process. Following Granger and Teräsvirta [27], Lin and Teräsvirta [48] solved the problem by devising a sequence of nested hypotheses. A similar hypotheses sequence has already been discussed in section 3.2, where it was used in the process of STR model specification. First, test the hypothesis (3.58) under the assumption of model (3.57) with k=3. If (3.58) is rejected, proceed by testing the hypothesis

$$H_{03} : \lambda_3 = 0,\ \delta_3 = 0 \tag{3.59}$$

against its obvious alternative hypothesis. While rejecting (3.59) signals the transition function F_3, the next hypothesis has to be tested if (3.59) is accepted, namely

$$H_{02} : \lambda_2 = 0,\ \delta_2 = 0 \mid \lambda_3 = 0,\ \delta_3 = 0. \tag{3.60}$$

If the hypothesis (3.60) is accepted, use F_1, otherwise use F_2. Testing the last hypothesis in the sequence (after accepting all of the previous hypotheses), namely

$$H_{01} : \lambda_1 = 0,\ \delta_1 = 0 \mid \lambda_2 = \lambda_3 = 0,\ \delta_2 = \delta_3 = 0, \tag{3.61}$$

is equivalent to using the original parameter constancy test against k=1.

Camacho [10] suggests a more general function to characterize the alternative to parameter constancy (the integer k is not bounded to the case of 1, 2 or 3). The description is given in chapter 4, since Camacho's work was done in the VAR setting.

3.4.4 Other misspecification tests

The LM test of no autoregressive conditional heteroscedasticity of Engle [19] and McLeod and Li [55], and the Lomnicki-Jarque-Berra test of the normal distribution of errors are performed in the same way as in the linear setting and are therefore not discussed here.

All 3 misspecification tests described in this section are designed as tests against specified parametric alternatives. In Granger and Teräsvirta [28], a comprehensive overview of tests used in nonlinear modeling can be found, including nonparametric tests.

3.5 Comparison with linear models

Several applied studies were conducted, where the comparison of linear regression models and STR models was in favor of the STR models. Most recently Teräsvirta, van Dijk and Medeiros [75] examined the forecasting accuracy of linear autoregressive and smooth transition autoregressive time series models for several macroeconomic variables of the G7 countries. The results show that the STAR models usually outperform the linear AR models. The authors also point out that careful specification is crucial in nonlinear time series modeling. Camacho [10] examined the nonlinear forecasting power of the composite index of leading indicators to predict both output growth and the business-cycle phases of the US economy. The obtained smooth transition vector autoregressive (STVAR) model was found to be superior to the linear VAR model with respect to the forecasting ability. The STVAR models are discussed in the next chapter. Weise [80] investigated whether monetary policy has asymmetric effects on output and prices using a nonlinear STR approach. He found out that output responds more strongly to negative shocks than to positive shocks in low growth states, while price level responds more strongly to negative shocks regardless of state. This again shows that working with a nonlinear model generates a result that is known in practice, but not achieved by linear models.

3.6 Appendix to chapter 3: Forms of the Lagrange multiplier test

Because of the importance of the Lagrange multiplier (LM) test for testing the null hypothesis of linearity, different forms of the test are studied in a separate section. Extensive discussion on this topic can be found in Greene [28], Harvey [32] and Davidson and MacKinnon [16]. The asymptotic properties of the maximum likelihood estimator and the definitions of the Likelihood ratio, Wald and Lagrange multiplier test are given in appendix A.

Suppose we would like to test if the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$ satisfies a set of q conditions or restrictions. If the restrictions are linear, they can be written as

$$R\theta = r, \tag{3.62}$$

where R is a full rank matrix with dimensions $q \times m$ and r is a vector of length q. To test such a hypothesis, three asymptotically equivalent test procedures are available, namely the Likelihood ratio, Wald and Lagrange multiplier test. In a more general formulation, nonlinear restrictions are also allowed. In this case, the restrictions under the null hypothesis take the form

$$H_0 : c(\theta) = r, \tag{3.63}$$

where $c(\theta) = (c_1(\theta), c_2(\theta), \ldots, c_q(\theta))'$ and the functions $c_j(\theta)$, $j = 1, 2, \ldots, q$, are continuously differentiable. The alternative hypothesis is

$$H_1 : c(\theta) \neq r. \tag{3.64}$$

Recall that the Lagrange multiplier test is used when the restricted maximum likelihood estimator $(\hat{\theta}_R)$ is easier to compute than the unrestricted estimator $(\hat{\theta}_U)$. A typical example is a nonlinear model, which becomes linear under the imposed restrictions. Intuitively, if the restrictions are valid, then the vector of partial derivatives $\frac{\partial \ln L(\theta)}{\partial \theta}$ (the score vector) evaluated at the restricted maximum likelihood estimator $\hat{\theta}_R$ will be close to zero. The Lagrange multiplier test statistic for testing null hypothesis (3.63) is of the form

$$LM = \left(\frac{\partial \ln L(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}_R} \right)' \left(I(\hat{\theta}_R) \right)^{-1} \left(\frac{\partial \ln L(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}_R} \right). \tag{3.65}$$

Under the null hypothesis, the LM test statistic is asymptotically χ^2-distributed with q degrees of freedom. For details, see appendix A. To make the notation shorter and easier to read, we shall write $\frac{\partial \ln L(\hat{\theta})}{\partial \hat{\theta}}$ instead of

$$\frac{\partial \ln L(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}},$$

thus using the same symbol for a variable and its value at a fixed point.

3.6.1 Restrictions involving only a subset of parameters

In practice, it is often the case that only a subset of parameters has to be tested. Let us partition the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$ into two subvectors,

$$\theta = \begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \end{pmatrix}, \tag{3.66}$$

such that the restrictions under the null hypothesis involve only the vector $\theta^{(2)}$ of length m_2. Thus, the null hypothesis is

$$H_0 : c(\theta^{(2)}) - r = 0. \tag{3.67}$$

The information matrix has to be partitioned accordingly as

$$I(\theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \tag{3.68}$$

where I_{11}, I_{12}, I_{21} and I_{22} are matrices of dimensions $m_1 \times m_1$, $m_1 \times m_2$, $m_2 \times m_1$ and $m_2 \times m_2$, respectively. Obviously,

$$\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R^{(1)}} = 0, \tag{3.69}$$

as the components of the vector $\theta^{(1)}$ do not enter the constraints under H_0. Note that the zero on the right-hand side of the previous equation stands for the m_1-dimensional vector of zeroes. By taking into consideration the formula for the partitioned inverse, the Lagrange multiplier test statistic

$$LM1 = \left(0' \quad \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R^{(2)}}\right)'\right) \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R^{(2)}} \end{pmatrix} \tag{3.70}$$

can be written as

$$LM2 = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R^{(2)}}\right)' \left(I_{22} - I_{21} I_{11}^{-1} I_{12}\right)^{-1} \frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R^{(2)}}. \tag{3.71}$$

It has to be emphasized that every block of the information matrix is computed at the restricted maximum likelihood estimator $\hat{\theta}_R$. Additionally, if the information matrix is block diagonal under the null hypothesis, that is $I_{12} = I_{21} = 0$, the test statistic LM2 simplifies to

$$LM3 = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R^{(2)}}\right)' \left(I_{22}\right)^{-1} \frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R^{(2)}}. \tag{3.72}$$

An important special case of restrictions involving only a subset of parameters, which is often used when testing the null hypothesis of linearity of a model, is

$$H_0 : \theta^{(2)} = 0. \tag{3.73}$$

3.6.2 LM test and nonlinear regression models

Generally, any regression model can be expressed as

$$y_t = g(x_t; \beta) + u_t, \quad t = 1, 2, \ldots, n. \tag{3.74}$$

We say that regression model (3.74) is nonlinear, if the function g is nonlinear in parameters (stacked in the parameter vector β). The vector of all parameters of the model is given by $\theta = (\beta', \sigma^2)'$. Let us denote the sum of squares that has to be minimized by

$$S(\beta) = \sum_{t=1}^{n} \Big(y_t - g(x_t; \beta)\Big)^2. \tag{3.75}$$

Provided that the residuals u_t, $t = 1, 2, \ldots, n$, are normally distributed with mean 0 and the covariance matrix equal to σ^2 times the identity matrix, the log-likelihood function can be computed as

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}S(\beta). \tag{3.76}$$

It follows from equation (3.76) that the maximum likelihood estimator of the parameter vector β and its nonlinear least squares estimator coincide. Both estimators are determined by solving the equation

$$\frac{\partial S(\beta)}{\partial\beta} = 0. \tag{3.77}$$

In order to estimate the information matrix, the following partial derivatives need to be computed:

$$\begin{aligned} \frac{\partial \ln L(\theta)}{\partial\beta} &= \frac{1}{\sigma^2}\sum_{t=1}^{n} h_t\Big(y_t - g(x_t;\beta)\Big) = \frac{1}{\sigma^2}\sum_{t=1}^{n} h_t u_t, \\ h_t &= \frac{\partial g(x_t;\beta)}{\partial\beta} = -\frac{\partial u_t}{\partial\beta}. \end{aligned} \tag{3.78}$$

After differentiating for the second time and taking into account the relation

$$E(y_t - g(x_t;\beta)) = E(u_t) = 0, \tag{3.79}$$

one obtains

$$E\Big(-\frac{\partial^2 \ln L(\theta)}{\partial\beta\partial\beta'}\Big) = \frac{1}{\sigma^2}\sum_{t=1}^{n} h_t h_t' \tag{3.80}$$

and

$$E\Big(-\frac{\partial^2 \ln L(\theta)}{\partial\beta\partial\sigma^2}\Big) = E\Big(\frac{1}{\sigma^4}\sum_{t=1}^{n} h_t u_t\Big) = 0. \tag{3.81}$$

We have just shown that the information matrix is block diagonal, which simplifies the Lagrange multiplier test statistic. As a direct consequence of Theorem A.7, the estimator $\hat{\beta}$ is asymptotically normally distributed with the asymptotic covariance matrix equal to $\sigma^2\big(\sum_{t=1}^{n} h_t h_t'\big)^{-1}$.

While applying the Lagrange multiplier test in the process of smooth transition regression modeling, we never impose restrictions on the parameter σ^2. It is therefore reasonable to construct the LM test for testing only the constraints involving the parameter vector β. Let $\hat{\beta}_R$ denote the restricted ML estimator of model (3.74) under the null hypothesis

$$H_0 : c(\beta) - r = 0. \tag{3.82}$$

Combining the results from equations (3.78) and (3.80) gives the test statistic

$$LM4 = \frac{1}{\hat{\sigma}_R^2}\left(\sum_{t=1}^{n} \hat{h}_t \hat{u}_t\right)'\left(\sum_{t=1}^{n} \hat{h}_t \hat{h}_t'\right)^{-1}\left(\sum_{t=1}^{n} \hat{h}_t \hat{u}_t\right) \tag{3.83}$$

for testing null hypothesis (3.82). The sign hat ($\hat{}$) above the variables is used to denote their values at the restricted estimator $\hat{\beta}_R$, for example

$$\hat{\sigma}_R^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2 = \frac{1}{n}\sum_{t=1}^{n} u_t^2(\hat{\beta}_R). \tag{3.84}$$

3.6.3 Using the coefficient of multiple correlation

One of the most popular forms of the Lagrange multiplier test statistic involves the coefficient of multiple correlation R^2 of a suitably chosen auxiliary regression. Recall that for the classical linear regression model

$$y_t = x_t'\beta + u_t, \quad t = 1, 2, \ldots, n, \tag{3.85}$$

or in the matrix form

$$y = X\beta + u, \tag{3.86}$$

R^2 is defined as the explained sum of squares divided by the total sum of squares (in the deviated form), namely

$$R^2 = \frac{\sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{n} (y_t - \bar{y})^2}. \tag{3.87}$$

Some algebraic manipulation yields

$$\begin{aligned} R^2 &= \frac{\sum_{t=1}^{n} \hat{y}_t^2 - n\bar{y}^2}{\sum_{t=1}^{n} y_t^2 - n\bar{y}^2} = \frac{\hat{\beta}'X'X\hat{\beta} - n\bar{y}^2}{y'y - n\bar{y}^2} \\ &= \frac{y'X(X'X)^{-1}X'y - n\bar{y}^2}{y'y - n\bar{y}^2} \\ &= \frac{\left(\sum_{t=1}^{n} x_t y_t\right)'\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}\left(\sum_{t=1}^{n} x_t y_t\right) - n\bar{y}^2}{\sum_{t=1}^{n} y_t^2 - n\bar{y}^2}. \end{aligned} \tag{3.88}$$

In particular, if $\bar{y} = 0$, then the previous expression simplifies to

$$R^2 = \frac{\left(\sum_{t=1}^{n} x_t y_t\right)' \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{n} x_t y_t\right)}{\sum_{t=1}^{n} y_t^2}. \tag{3.89}$$

The Lagrange multiplier test statistic LM4 defined by equation (3.83) can be rewritten as

$$LM5 = \frac{n}{\sum_{t=1}^{n} \hat{u}_t^2} \left(\sum_{t=1}^{n} \hat{h}_t \hat{u}_t\right)' \left(\sum_{t=1}^{n} \hat{h}_t \hat{h}_t'\right)^{-1} \left(\sum_{t=1}^{n} \hat{h}_t \hat{u}_t\right). \tag{3.90}$$

The comparison of LM5 and the last form of the multiple correlation coefficient R^2 says that the LM test for testing the restrictions

$$H_0 : c(\beta) - r = 0 \tag{3.91}$$

imposed on the parameters of the model

$$y_t = g(x_t; \beta) + u_t, \quad t = 1, 2, \ldots, n, \tag{3.92}$$

can also be performed with the help of the statistic

$$LM6 = nR^2, \tag{3.93}$$

where R^2 is the squared multiple correlation coefficient obtained from the auxiliary regression of $\hat{u}_t$ on $\hat{h}_t$. Note that the conclusion holds only if

$$\bar{\hat{u}} = \frac{1}{n} \sum_{t=1}^{n} \hat{u}_t = 0.$$

This is the case when nonlinear regression model (3.92) contains a constant term, although the derivation still holds asymptotically even if the constant term is not present in the model.

3.6.4 LM test and linear regression models

As a special case of model (3.92), we consider a linear regression model

$$y_t = x_t' \beta + u_t, \quad t = 1, 2, \ldots, n, \tag{3.94}$$

while the restrictions under the null hypothesis remain the same. Evidently, $h_t = \frac{\partial(x_t'\beta)}{\partial\beta} = x_t$ in this setting. Transforming the LM5 test statistic given by equation (3.90) now yields

$$LM7 = \frac{1}{\hat{\sigma}_R^2}\left(\sum_{t=1}^{n} x_t\hat{u}_t\right)'\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}\left(\sum_{t=1}^{n} x_t\hat{u}_t\right). \tag{3.95}$$

On the other hand, the LM7 statistic is still of the form nR^2, but R^2 stands for the squared multiple correlation coefficient obtained from the auxiliary regression of $\hat{u}_t$ on x_t. Note that $\hat{u}_t$ denotes the residuals of the restricted model.

In the STR modeling process, one often has to test the null hypothesis that a subset of parameters in an auxiliary linear regression model of the form (3.94) is equal to zero. Therefore, the parameter vector $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ is partitioned into two subvectors,

$$\beta = \begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix}, \tag{3.96}$$

similarly as in subsection 3.6.1. The null hypothesis reads

$$H_0 : \beta^{(2)} = 0, \tag{3.97}$$

where the vector $\beta^{(2)}$ is of length k_2. The vector x_t is partitioned into a k_1-vector z_t and a k_2-vector w_t and model (3.94) becomes

$$y_t = z_t'\beta^{(1)} + w_t'\beta^{(2)} + u_t, \quad t = 1, 2, \ldots, n. \tag{3.98}$$

To derive the Lagrange multiplier test statistic, we only have to replace $g(x_t;\beta)$ with $z_t'\beta^{(1)}+w_t'\beta^{(2)}$ in subsection 3.6.2 and then modify the derivation accordingly. The results of subsection 3.6.1 are also taken into account. Obviously,

$$\begin{aligned} \frac{\partial \ln L(\theta)}{\partial \beta^{(1)}} &= \frac{1}{\sigma^2}\sum_{t=1}^{n} z_t u_t, \\ \frac{\partial \ln L(\theta)}{\partial \beta^{(2)}} &= \frac{1}{\sigma^2}\sum_{t=1}^{n} w_t u_t, \end{aligned} \tag{3.99}$$

and

$$
\begin{aligned}
I_{11} &= E\left(-\frac{\partial^2 \ln L(\theta)}{\partial\beta^{(1)}\partial\beta^{(1)\prime}}\right) = \frac{1}{\sigma^2}\sum_{t=1}^{n} z_t z_t', \\
I_{22} &= E\left(-\frac{\partial^2 \ln L(\theta)}{\partial\beta^{(2)}\partial\beta^{(2)\prime}}\right) = \frac{1}{\sigma^2}\sum_{t=1}^{n} w_t w_t', \\
I_{12} = I_{21}' &= E\left(-\frac{\partial^2 \ln L(\theta)}{\partial\beta^{(1)}\partial\beta^{(2)\prime}}\right) = \frac{1}{\sigma^2}\sum_{t=1}^{n} z_t w_t'. \qquad (3.100)
\end{aligned}
$$

The LM2 test statistic (3.71) now reads

$$
LM8 = \frac{1}{\hat{\sigma}_R^2}\left(\sum_{t=1}^{n} w_t\hat{u}_t\right)'\left(M_{22} - M_{21}M_{11}^{-1}M_{12}\right)^{-1}\left(\sum_{t=1}^{n} w_t\hat{u}_t\right), \qquad (3.101)
$$

where $M_{11} = \sum_{t=1}^{n} z_t z_t'$, $M_{22} = \sum_{t=1}^{n} w_t w_t'$, $M_{12} = M_{21}' = \sum_{t=1}^{n} z_t w_t'$, $\hat{u}_t$ is the residual estimated under the null hypothesis and the restricted estimator $\hat{\sigma}_R^2$ is equal to $\frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2$.

3.6.5 Modified LM test

Suppose that we would like to test null hypothesis (3.97) for model (3.98) with an LM test. If the number of the elements of x_t is large when compared to the sample size, the asymptotic χ^2 distribution is likely to be a poor approximation for the actual small sample distribution of the LM test statistic. In this case, an F - approximation of the test statistic works much better, as the empirical size of the test remains close to the nominal size while the power is good.

Let us denote the vector of the estimated OLS residuals for the model

$$
y_t = x_t'\beta + u_t = z_t'\beta^{(1)} + w_t'\beta^{(2)} + u_t \qquad (3.102)
$$

by $\hat{u}$, the vector of the estimated residuals of the restricted model

$$
y_t = z_t'\beta^{(1)} + u_{0t} \qquad (3.103)
$$

by $\hat{u}_0$ and the residuals obtained when regressing $\hat{u}_{0t}$ on x_t by $\hat{\varepsilon}$. The corresponding sums of squares are denoted by $SSR = \hat{u}'\hat{u}$, $SSR_0 = \hat{u}_0'\hat{u}_0$ and $SSR_\varepsilon = \hat{\varepsilon}'\hat{\varepsilon}$, respectively. It follows from subsection 3.6.4 that the LM test statistic for testing null hypothesis (3.97) is nR^2, where R^2 denotes the squared multiple correlation coefficient from regressing $\hat{u}_{0t}$ on x_t. In our notation,

$$R^2 = \frac{SSR_0 - SSR_\varepsilon}{SSR_0}. \tag{3.104}$$

However, $SSR_\varepsilon = SSR$, which can be shown as follows. By using the standard notation $M = I - X(X'X)^{-1}X'$ and taking into account that $MX = 0$, one obtains

$$\hat{\varepsilon} = M\hat{u}_0 = M\left(y - X\begin{pmatrix} \hat{\beta}^{(1)} \\ 0 \end{pmatrix}\right) = My = \hat{u}. \tag{3.105}$$

Thus, the residuals from the regression of $\hat{u}_{0t}$ on x_t coincide with the residuals obtained when regressing y_t on x_t. The Lagrange multiplier test statistic can be written as

$$LM9 = nR^2 = n\frac{SSR_0 - SSR}{SSR_0} = n\left(1 - \frac{SSR}{SSR_0}\right). \tag{3.106}$$

The standard F-statistic for testing null hypothesis (3.97) is

$$F = \frac{(SSR_0 - SSR)/k_2}{SSR/(n-k)}. \tag{3.107}$$

It is easy to verify that the equation

$$LM9 = \frac{n\frac{k_2}{n-k}F}{\frac{k_2}{n-k}F + 1} \tag{3.108}$$

holds. Consequently, the LM9 statistic is a monotonous function of the F-statistic and the Lagrange multiplier principle leads to the F-test.

Chapter 4

Systems of equations

An important problem in the identification process of economic systems is usually related to the question whether the model can be kept linear or whether nonlinear features are so dominant that they must be considered in the specification. From recent studies of univariate models one has learned that there is much to be gained by allowing a nonlinear specification. Representations of asymmetric reactions, structural changes and other phenomena of economic development can be fruitfully investigated by nonlinear modeling techniques. As many issues in economics require the specification of several relationships, techniques to handle nonlinear features in systems are required. Only during the recent years such methods have appeared in the literature. Most of the work has been done in the nonlinear VAR framework.

Anderson and Vahid [2] devised a procedure for detecting common nonlinear components in a multivariate system of variables. The common nonlinearities approach is based on the canonical correlations technique and can help us interpret the relationships between different economic variables. The specification and estimation of the system of equations is also simplified, since the existence of common nonlinearities reduces the dimension of nonlinear components in the system and enables parsimony. This is particulary important in empirical investigations involving economic time series of shorter length. Namely, most of the macroeconomic indicators are published on a quarterly basis.

Weise [80], van Dijk [78] and Camacho [10] extended the STR modeling approach developed by Teräsvirta and coworkers to vector autoregressive models of smooth transition. Their STR specification is limited to the case where the transition between different parameter regimes is governed by the same transition variable and the same type of transition function in every equation of the system. They argue that since the economic practice imposes common nonlinear features, all equations share the same switching regime. But this argument is not convincing, since such a conclusion cannot be derived from economic theory, while applied econometric studies analyzing nonlinear systems are scarce. For this reason we shall try to extend their approach by allowing different smooth transition functional forms in different equations. The proposed augmented specification procedure is explained in section 4.3.

4.1 Testing multiple equation systems for common nonlinear components

Anderson and Vahid [2] describe a generalized method of moment test for common nonlinear components in multiple time series. The number of nonlinear functions that need to be estimated can be reduced if the system contains common nonlinear components. The basic idea behind their work is to detect all linear combinations of the (possibly nonlinear) variables that do not exhibit nonlinear properties. The number of such linear combinations determines the number of common nonlinear components. The usual statistic and econometric tests are adapted to meet the needs of multivariate systems. The canonical correlations procedure is used to obtain the estimates of the linear combinations without the nonlinear properties. The paper discusses the tests for which the alternative is specified and also those for which the alternative is not specified.

4.1.1 Definition of common nonlinearity

Let us first explain the notion of common nonlinearity as defined in [2]. Suppose that the conditional mean of the i - th component of an n - dimensional vector y_t given a k - dimensional vector x_t can be written as

$$E(y_{it}|x_t) = \beta_i^{'} x_t + \psi_i(x_t, \theta_i), \quad i = 1, \ldots, n, \tag{4.1}$$

where the function ψ_i is nonlinear in x_t and possibly also in the parameter θ_i. If one can find $s < n$ linearly independent linear combinations of the components of the vector y_t with a linear conditional mean, than there exists an $n \times s$ matrix A with a full column rank for which the equation

$$A'\psi(x_t, \theta) = 0$$

holds. ψ and θ stand for the vectors $(\psi_1, \psi_2, \ldots, \psi_n)'$ and $(\theta_1, \theta_2, \ldots, \theta_n)'$, respectively. Obviously, A is not unique. As the matrix $A \cdot H$, where H is any $s \times s$ nonsingular matrix, also satisfies the previous equation, the matrix A can be normalized without loss of generality. A particularly useful normalization contains an $s \times s$ identity matrix as the first block. The matrix ψ is partitioned analogously:

$$A = \begin{bmatrix} I_s \\ A^{**} \end{bmatrix}, \quad \psi(x_t, \theta) = \begin{bmatrix} \psi^*(x_t, \theta) \\ \psi^{**}(x_t, \theta) \end{bmatrix}.$$

From $A'\psi(x_t, \theta) = 0$ follows $\psi^*(x_t, \theta) = -A^{**'}\psi^{**}(x_t, \theta)$ and

$$\psi(x_t, \theta) = \begin{bmatrix} -A^{**'} \\ I_{n-s} \end{bmatrix} \psi^{**}(x_t, \theta).$$

Finally, we can eliminate s nonlinear components and therefore write the conditional expectation of the vector y_t in terms of only n-s nonlinear components:

$$E(y_t|x_t) = Bx_t + A^{\perp}\psi^{**}(x_t, \theta), \tag{4.2}$$

with $A^{\perp} = \begin{bmatrix} -A^{**'} \\ I_{n-s} \end{bmatrix}$ and B equal to the $n \times k$ matrix of stacked vectors β_i', $i = 1, \ldots, n$.

Definition 4.1. *We say that the system (4.1) has n - s common nonlinear components when it is possible to rewrite (4.1) in the form (4.2) and if s is the largest integer with this property.*

4.1.2 Canonical correlations

It is often the case that when given two large groups of variables, one would like to study the interrelations. The technique of canonical correlations enables such between - group comparisons. An extensive discussion of canonical correlations can be found in a book by T.W. Anderson [3].

Let z denote an n - dimensional vector with the covariance matrix Σ, which is assumed to be positive definite, partitioned into subvectors z_1 and z_2,

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

where z_1 is an n_1 - dimensional vector and z_2 is an n_2 - dimensional vector with $n_1 \leq n_2$. By applying the same partitioning to the covariance matrix we can write it in the form

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{4.3}$$

Additionally, assume that $E(z) = 0$. Our task is to determine real vectors α with n_1 components and γ with n_2 components such that the correlation between the variables $u = \alpha^{'} z_1$ and $v = \gamma^{'} z_2$ is maximal. As multiplication of α or γ by a positive number does not change the correlation coefficient, it is sensible to perform the following normalization:

$$1 = E(u^2) = E(\alpha^{'} z_1 z_1^{'} \alpha) = \alpha^{'} \Sigma_{11} \alpha, \tag{4.4}$$

$$1 = E(v^2) = E(\gamma^{'} z_2 z_2^{'} \gamma) = \gamma^{'} \Sigma_{22} \gamma. \tag{4.5}$$

Evidently,

$$Corr(u, v) = E(uv) = E(\alpha^{'} z_1 z_2^{'} \gamma) = \alpha^{'} \Sigma_{12} \gamma. \tag{4.6}$$

Thus, we have to find the vectors α and γ that maximize (4.6) subject to (4.4) and (4.5). If we set

$$\psi = \alpha^{'} \Sigma_{12} \gamma - \frac{1}{2} \lambda (\alpha^{'} \Sigma_{11} \alpha - 1) - \frac{1}{2} \mu (\gamma^{'} \Sigma_{22} \gamma - 1), \tag{4.7}$$

where λ and μ are the Lagrange multipliers, then (after some algebraic manipulation) the Lagrange multiplier technique yields a system of equations

$$\begin{pmatrix} -\lambda \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda \Sigma_{22} \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = 0. \tag{4.8}$$

The previous system has a nontrivial solution if and only if

$$\begin{vmatrix} -\lambda \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda \Sigma_{22} \end{vmatrix} = 0. \tag{4.9}$$

The left - hand side of the above expression is a polynomial in λ of degree n. Let us write its n roots in a decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n. \tag{4.10}$$

Definition 4.2 ([3], Definition 12.2.1). *Let $z = \binom{z_1}{z_2}$, where z_1 has n_1 components and z_2 has n_2 $(= n - n_1 \geq n_1)$ components. The* r*th pair of canonical variates are the pair of linear combinations $u_r = \alpha_r' z_1$ and $v_r = \gamma_r' z_2$, each of unit variance and uncorrelated with the first r–1 pairs of canonical variates, having maximum correlation. The correlation is the* r*th canonical correlation.*

Theorem 4.3 ([3], Theorem 12.2.1). *Let $z = \binom{z_1}{z_2}$ be a random vector with covariance matrix Σ. The rth canonical correlation λ_r between z_1 and z_2 is the rth largest root of equation (4.9). The vectors α_r and γ_r defining the rth pair of canonical variates $u_r = \alpha_r' z_1$ and $v_r = \gamma_r' z_2$ satisfy (4.8) for $\lambda = \lambda_r$, and (4.4) and (4.5).*

For the proof, see [3].

4.1.3 The common nonlinearities test

Since the common nonlinearities test is deduced from the test of overidentifying restrictions in the generalized method of moments (GMM) framework, it has an asymptotic χ^2 - distribution provided that the regularity conditions described below hold. A thorough discussion of necessary and sufficient conditions is given in Wooldridge [83]. Regularity conditions are:

1. The variables are essentially stationary and weekly dependent.
   The stochastic process $\{x_t : t \in \mathbb{N}\}$ is *essentially stationary*, if the set $\{E(x_t^2) : t \in \mathbb{N}\}$ is bounded. When, in addition,

$$\sigma_T^2 \equiv Var\left(\sum_{t=1}^{T} x_t\right) = O(T), \qquad \sigma_T^{-2} = O(T^{-1}), \tag{4.11}$$

   and the central limit theorem holds, namely,

$$\frac{\sum_{t=1}^{T}\big(x_t - E(x_t)\big)}{\sigma_T} \xrightarrow{d} N(0,1), \tag{4.12}$$

   then the process is *weekly dependent.*

2. The assumptions GMM.1 to GMM.10 (as defined in Wooldridge) hold. These conditions guarantee that the GMM estimator exists, is consistent and asymptotically normally distributed.

Suppose that the null hypothesis of linearity is rejected for every component of the n - vector y_t. In this case, the conditional mean of y_{it}, $i = 1, \ldots, n$, is of the form (4.1). If there is a linear combination $\alpha_1^{'} y_t$ of the components of the vector y_t with a linear conditional mean, than the number of nonlinear components in the system can be reduced. A reduction of this kind requires

$$E\Big((\alpha_1^{'} y_t^{\dagger}) \otimes w_t^{\dagger}\Big) = 0, \tag{4.13}$$

where $\otimes$ denotes the Kronecker product. Note that $y_t^{\dagger}$ has had the part linear in x_t removed. $w_t^{\dagger}$ stands for the m - dimensional vector of nonlinear regressors, as applied in the univariate setting after substituting the transition function with its Taylor approximation. Exact definition of $w_t^{\dagger}$ in case of a smooth transition vector autoregressive model is given on page 63. Because of the structure of equation (4.13), we can use the generalized method of moments (GMM) to estimate the vector α_1. By replacing the moment condition with the corresponding sample mean $\frac{1}{T}\sum_{t=1}^{T}\Big((\alpha_1^{'} y_t^{\dagger}) \otimes w_t^{\dagger}\Big)$, we can construct the objective function (see Greene [28], p. 538):

$$\begin{aligned} Q &= \frac{1}{T}\sum_{t=1}^{T}\Big((\alpha_1^{'} y_t^{\dagger}) \otimes w_t^{\dagger}\Big)^{'} \times \hat{V}_{1T}^{-1} \times \frac{1}{T}\sum_{t=1}^{T}\Big((\alpha_1^{'} y_t^{\dagger}) \otimes w_t^{\dagger}\Big) \qquad (4.14) \\ &= \frac{1}{T^2}\alpha_1^{'} Y^{'} W \hat{V}_{1T}^{-1} W^{'} Y \alpha_1. \end{aligned}$$

The optimal GMM estimator of α_1 minimizes the above objective function. While Y and W denote the matrices with stacked vectors $y_t^{\dagger}$ and $w_t^{\dagger}$ as rows, $\hat{V}_{1T}$ stands for the matrix that complies with the condition

$$plim \hat{V}_{1T} = \lim_{T \to \infty} E(T^{-1}(W^{'} Y \alpha_1 \alpha_1^{'} Y^{'} W)). \tag{4.15}$$

The question we have to answer is how to estimate the rank and the basis of the space of vectors that satisfy equation (4.13). Let the matrix α consist of s such vectors, $\alpha_1, \alpha_2, \ldots, \alpha_s$, written as columns. If the operator $vec(A)$ concatenates the columns of the given matrix A into one column vector, then the objective function for estimating the matrix α with the generalized method of moments is of the form

$$Q = \frac{1}{T^2} vec^{'}(Y\alpha)(I_s \otimes W)\hat{V}_T^{-1}(I_s \otimes W^{'})vec(Y\alpha),$$

where

$$plim\hat{V}_T = \lim_{T\to\infty} E\left[\frac{1}{T}(I_s \otimes W^{'})vec(Y\alpha)vec^{'}(Y\alpha)(I_s \otimes W)\right].$$

Under the assumptions of no serial correlation and no heteroscedasticity,

$$\hat{V}_T = \hat{\Sigma} \otimes \frac{W^{'}W}{T}$$

turns out to be a suitable choice, with $\hat{\Sigma}$ equal to a consistent estimate of $E(\alpha^{'}y_t^{\dagger}y_t^{\dagger'}\alpha)$.

Anderson and Vahid propose to perform the test in several steps:

1. Calculate $\hat{\Sigma}$;

2. Determine the rank of α in a loop. Set s to 0. While overidentifying restrictions are not rejected, do the following:

   a. increase s by 1,

   b. perform the GMM estimation,

   c. test the overidentifying restrictions.

The identifying restrictions can be carried out by setting $\hat{\Sigma}$ equal to the identity matrix, which also simplifies the procedure, as shown by the next lemma.

Lemma 4.4 ([2], Lemma 2.2). *Under the normalization*

$$(1/T)\alpha^{'}Y^{'}Y\alpha = I_s,$$

1. *the GMM estimators of columns of α are the canonical coefficient vectors of $y_t^{\dagger}$ corresponding to $\{\lambda_i^2,\ i = 1, \ldots, s\}$, the s smallest estimated squared canonical correlations between $y_t^{\dagger}$ and $w_t^{\dagger}$.*

2. *The test statistic for the overidentifying restrictions is $T\sum_{i=1}^{s}\lambda_i^2$, which has the same asymptotic distribution as the statistic $-T\sum_{i=1}^{s}\ln(1-\lambda_i^2)$. (This latter statistic is the standard likelihood ratio test of the null that the s smallest canonical correlations are zero under the assumption of normality.)*

The proof of the lemma is given in [2]. As already mentioned, the regularity conditions guarantee the asymptotic χ^2 distribution of the common nonlinearities test statistic. The degree of freedom is equal to the number of overidentifying restrictions, in our case $(m-n)s+s^2$. Observe that the number of moment conditions in

$$E\Big((\alpha_j' y_t^\dagger) \otimes w_t^\dagger\Big) = 0, \quad j = 1, 2, \ldots, s, \tag{4.16}$$

is equal to ms, whereas the number of parameters (i.e. components of the matrix α) to be estimated is $ns - s^2$, since s^2 parameters are determined by normalization. Therefore, the number of overidentifying restrictions is equal to $ms - (ns - s^2)$.

Testing for common STAR nonlinearities

Definition 4.5. *A smooth transition vector autoregressive model, shortly STVAR(p), is a model of the form*

$$y_t = A_0 + A_1(L)y_t + G(s_t)[B_0 + B_1(L)y_t] + \varepsilon_t, \tag{4.17}$$

where y_t is an n - vector time series, A_0 and B_0 are n - vectors of constants, $A_1(L)$ and $B_1(L)$ are pth order matrix polynomials in the lag operator with $A_1(0) = B_1(0) = 0$, ε_t is an $n \times 1$ i.i.d. $(0, \Sigma)$ sequence, G is an $n \times n$ diagonal transition matrix with a typical diagonal entry $G_i(s_{it})$, and s_{it} is one of the np lagged regressors in $ylags = (y_{t-1}', y_{t-2}', \ldots, y_{t-p}')'$. The specification that each transition function $G_i(s_{it})$ is a logistic function of the form

$$G_i(s_{it}) = (1 + \exp[-\gamma_i(s_{it} - c_i)])^{-1},$$

where $\gamma > 0$, leads to the logistic STVAR model.

In the study of univariate LSTAR models by Lukkonen et al. [51] and Teräsvirta [73], the models with $s_{it} = y_{it-d}$ are discussed. After the logistic transition function is replaced by its third order Taylor approximation around $\gamma = 0$, the univariate linearity test is developed. The procedure is explained in detail in section 3.1. Testing a single variable in the multivariate setting is similar. The null hypothesis of linearity, i.e.

$$H_0 : \beta_{2j} = \beta_{3j} = \beta_{4j} = 0, \quad j = 1, \ldots, np,$$

based on the auxiliary regression

$$y_{it} = \beta_0 + \beta_1^{'}(ylags_t) + \beta_2^{'}(ylags_t \cdot s_{it}) + \beta_3^{'}(ylags_t \cdot s_{it}^2) + \beta_4^{'}(ylags_t \cdot s_{it}^3) + v_{it},$$

is tested against the alternative hypothesis of a STAR model. If the previously described regularity conditions hold true, then the test statistic is asymptotically χ^2-distributed.

Suppose that in the smoooth transition vector autoregressive setting the univariate LSTAR test is applied to $y_{1t}, y_{2t}, \ldots, y_{nt}$, every time with the same transition variable s_t. If the test rejects the null hypothesis of linearity for at least two variables, say y_{kt} and y_{lt}, it is possible that the variables share a common nonlinearity. In this case one can find at least one n - dimensional vector α for which the condition (4.13) is fulfilled:

$$E\Big((\alpha^{'} y_t^{\dagger}) \otimes w_t^{\dagger}\Big) = 0, \tag{4.18}$$

with $w_t^{\dagger}$ equal to $((ylags_t \cdot s_t)^{'}, (ylags_t \cdot s_t^2)^{'}, (ylags_t \cdot s_t^3)^{'})^{'}$. The sign † indicates that the influence of the linear terms, namely the constant and the $ylags$, has been removed from the vector w_t by regressing it on the constant and the components of the vector $ylags$. Taking into account the results from Lemma 4.4, the test statistic for the null hypothesis that there are at least s linearly independent linear combinations of the components of the vector y_t with a linear conditional mean is of the form

$$J = T \sum_{i=1}^{s} \lambda_i^2. \tag{4.19}$$

The λ_i are the estimated canonical correlations between the variable groups $y_t^{\dagger}$ and $w_t^{\dagger}$. Obviously, if there are s independent linear combinations with a linear mean in a model where all n dependent variables are nonlinear in mean, then there are n-s common nonlinear components. The null hypothesis is rejected when there are more than n-s common nonlinear components. Provided that the described regularity conditions hold, the test statistic J has an asymptotic χ^2-distribution with $(3p-1)ns + s^2$ degrees of freedom.

Finite sample properties of the test

Anderson and Vahid conducted several Monte Carlo experiments in order to evaluate the properties of the test when STAR nonlinearities are present.

They restricted their attention to the bivariate and trivariate case with 0, 1 or 2 and 0, 1, 2 or 3 nonlinearities of the LSTAR type with the same known transition variable in the true data generating process (DGP), respectively.

The simulations with sample sizes of 100 and 300 were preformed in each case. The bivariate LSTAR test performed well, especially in the sample size of 300, whereas the trivariate LSTAR test had difficulty distinguishing between the cases of 1, 2 and 3 common nonlinear components. It has to be emphasized that the chosen nonlinear data generating processes were close to linear because of the small values of the γ parameters. When performed on DGPs with higher γ values, the tests performed substantially better.

Applications

In their paper, Anderson and Vahid apply the common nonlinearities approach for modeling business cycles in Canada and United States, and for developing a nonlinear model of real US aggregates. Both nonlinear models are shown to be superior to their linear counterparts. It is well known that the business cycles are asymmetric insofar as the recessions are more pronounced but last only a short period of time, whereas the expansions are usually long - lasting but mild. For this reason the linear VAR models are not appropriate for modeling business cycles. Anderson and Vahid propose a smooth transition vector autoregressive model to be used instead. It turns out that one common nonlinear component can account for the asymmetries of business cycles in Canada and US. In the second empirical investigation, the linear real business cycle model proposed by King et al. [43] is reexamined and the hypothesis of linearity is rejected in favor of a logistic smooth transition vector autoregressive model. We have to note that both empirical nonlinear models exhibit potential for forecasting.

4.2 A smooth transition approach to vector autoregressive models

Whereas there has been extensive research in the field of univariate nonlinear modeling, the statistical theory of multivariate nonlinear modeling has yet to be developed. The first attempts at extending nonlinear smooth transition regression techniques to a multivariate setting can be found in Weise [80], van Dijk [78] and Camacho [10]. Similarly, multivariate Markov - switching models are treated in Krolzig [46] and multivariate threshold models in Tsay [77].

Weise [80], van Dijk [78] and Camacho [10] extended the univariate STR modeling approach developed by Teräsvirta and coworkers to vector autoregressive models of smooth transition. Their STR specification is limited to the case where the transition between different parameter regimes is governed by the same transition variable and the same type of transition function in every equation of the system. They argue that since the economic practice imposes common nonlinear features, all equations share the same switching regime. After using a three - equation linear structural model as a starting point, Weise [80] develops and analyzes the obtained reduced form, which is given in a form of a vector autoregressive model. Van Dijk [78] applies the STVAR modeling approach to study the intraday spots and futures prices of the FTSE100 index, whereas Camacho [10] examines the nonlinear forecasting power of the composite index of leading indicators to predict both output growth and the business - cycle phases of the US economy. Since all three studies are similar, while the most comprehensive description of the methodological approach is given by Camacho [10], we shall start with a short review of his work.

4.2.1 Specification and estimation

Camacho [10] considers a 2 - dimensional smooth transition vector autoregressive model

$$\begin{aligned} y_t &= \varphi_y' X_t + (\theta_y' X_t) G_y(s_{yt}) + u_{yt} \\ x_t &= \varphi_x' X_t + (\theta_x' X_t) G_x(s_{xt}) + u_{xt}, \end{aligned} \tag{4.20}$$

where $X_t = (1, y_{t-1}, x_{t-1}, \ldots, y_{t-p}, x_{t-p})' = (1, \tilde{X}_t')'$, $\varphi_y, \varphi_x, \theta_y, \theta_x$ are the corresponding parameter vectors and $U_t = (u_{yt}, u_{xt})' \sim N(0, \Omega)$ is a vector series of serially uncorrelated errors. The difference $D_{it} = s_{it} - c_i$, $i = x, y$, in the exponent of the transition function G_i is called *the switching expression*. The letters y_t and x_t are used for the two variables in the autoregressive system, since the smooth transition approach is applied to the rate of growth of US GDP and the rate of growth of the US composite index of leading indicators, respectively. The discussion is restricted to the case of $s_{xt} = s_{yt}$ and $G_x = G_y$, where the same transition variable and the same transition function is specified in both equations.

After the linear VAR has been specified, the linearity test is applied. The problems with nuisance parameters are solved with suitable Taylor series expansions, as usually. The auxiliary regression to be performed in case the transition variable s_t belongs to X_t is

$$\begin{aligned} y_t &= \eta_{y0}^{'} X_t + \sum_{h=1}^{3} \eta_{yh}^{'} \tilde{X}_t s_t^h + v_{yt} \\ x_t &= \eta_{x0}^{'} X_t + \sum_{h=1}^{3} \eta_{xh}^{'} \tilde{X}_t s_t^h + v_{xt} \end{aligned} \tag{4.21}$$

and the null hypothesis of linearity reads as

$$H_0 : \eta_{i1} = \eta_{i2} = \eta_{i3} = 0, \quad i = x, y. \tag{4.22}$$

Consequently, the null hypothesis can be tested with the Lagrange multiplier test discussed in the appendix to chapter three.

If the null hypothesis of linearity is rejected in favor of the alternative smooth transition vector autoregressive model, one has to decide which transition function to use. The decision is based on the sequence of nested hypotheses tests described in section 3.2.2. The parameters of the specified model are estimated with the maximum likelihood estimator under the assumption of normally distributed errors:

$$U_t = (u_{yx}, u_{xt})^{'} \sim N(0, \Omega). \tag{4.23}$$

4.2.2 Testing the model adequacy

As proposed by Eitrheim and Teräsvirta [18], three tests are performed in order to check for the adequacy of the estimated model, namely the Serially independent errors test (SI test), the Parameter constancy test (PC test) and the No remaining nonlinearity test (NRN test). The multivariate generalizations of the three test as developed by Camacho are discussed below.

Serially independent errors test

The serially independent errors test allows for the autocorrelation of residuals under the alternative hypothesis, i.e.

$$Y_t = F(X_t, \Psi) + U_t, \tag{4.24}$$

with $U_t = (u_{yt}, u_{xt})^{'}$, $Y_t = (y_t, x_t)^{'}$, $F(X_t, \Psi) = (F_y(X_t, \Psi_y), F_x(X_t, \Psi_x))^{'}$, $F_i(X_t, \Psi_i) = \varphi_i^{'} X_t + (\theta_i^{'} X_t) G_i(s_{it})$ and $\Psi = (\Psi_y^{'}, \Psi_x^{'})^{'}$. The vector

$$\Psi_i = (\varphi_i^{'}, \theta_i^{'}, \gamma_i, c_i)^{'}, \quad i = y, x, \tag{4.25}$$

contains all of the unknown parameters from $F_i(X_t, \Psi_i)$. Under the assumption of AR(r) errors, the vector U_t takes the form

$$U_t = \Lambda(L) U_t + \zeta_t, \quad \zeta_t \sim N(0, \Gamma), \tag{4.26}$$

where ζ_t is serially independent and

$$\Lambda(L) = \Lambda_1 L + \ldots + \Lambda_r L^r \tag{4.27}$$

is a 2×2 matrix polynomial of order r in the lag operator L that can be represented by a $2 \times 2r$ matrix

$$\widetilde{\Phi} = \begin{pmatrix} \Phi'_{yy} & \Phi'_{yx} \\ \Phi'_{xy} & \Phi'_{xx} \end{pmatrix}. \tag{4.28}$$

Note that Φ'_{ij}, $i, j \in \{x, y\}$, are row vectors of length r defined in an obvious way. In other words, if

$$V_t = (u_{y,t-1}, \ldots, u_{y,t-r}, u_{x,t-1} \ldots, u_{x,t-r})^{'}, \tag{4.29}$$

then the equation

$$\Lambda(L) U_t = \widetilde{\Phi} V_t \tag{4.30}$$

holds. Since the operation *vec* stacks the columns of a given matrix into a long column vector, the vector Φ defined by $\Phi = vec(\widetilde{\Phi}^{'}) = (\Phi'_{yy}, \Phi'_{yx}, \Phi'_{xy}, \Phi'_{xx})^{'}$ contains both rows of the matrix $\widetilde{\Phi}$ concatenated into a vector of length 4r. All parameters of the model that have to be estimated are contained in the vector $\vartheta = (\Psi^{'}, \Phi^{'})^{'}$. The null hypothesis of no serial correlation of errors can be tested by $H_0 : \Phi = 0$ with the Lagrange multiplier test. As the vector Φ contains 4r elements, the test statistics is asymptotically χ^2-distributed with 4r degrees of freedom. The standard Lagrange multiplier test statistics can be written as

$$LM = \frac{1}{T} m_{\Phi}^{'} \left(M_{\Phi\Phi} - M_{\Phi\Psi} (M_{\Psi\Psi})^{-1} M_{\Phi\Psi}^{'} \right)^{-1} m_{\Phi}. \tag{4.31}$$

While m_Φ is a vector of partial derivatives of the score, the matrices M are blocks of the partitioned information matrix.

Camacho simplifies the test statistics in the following way. Pre-multiplying equation (4.24) by $(I - \Lambda(L))$ gives

$$(I - \Lambda(L))Y_t = (I - \Lambda(L))F(X_t, \Psi) + \zeta_t, \tag{4.32}$$

with the likelihood function

$$L_t = \frac{1}{(2\pi)|\Gamma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\zeta_t'\Gamma^{-1}\zeta_t\right) \tag{4.33}$$

and the score

$$\begin{aligned} l_t &= C - \frac{1}{2}\ln|\Gamma| - \frac{1}{2}\left(\zeta_t'\Gamma^{-1}\zeta_t\right) \\ &= C - \frac{1}{2}\ln|\Gamma| - \frac{1}{2}\left(\zeta_{yt}^2\Gamma^{yy} + 2\zeta_{yt}\zeta_{xt}\Gamma^{yx} + \zeta_{xt}^2\Gamma^{xx}\right). \end{aligned} \tag{4.34}$$

Note that Γ^{ij}, $i,j \in \{x,y\}$, are blocks of the matrix Γ^{-1} and the vector ζ_t is partitioned in an analogous way. Let us denote by the underbar of an expression its maximum likelihood estimate under the null hypothesis. Taking into account equations (4.34) and (4.26), one can express the LM test statistic in terms of known quantities. From equation

$$\zeta_{it} = u_{it} - (\phi_{iy}', \phi_{ix}')(v_{yt}', v_{xt}')', \quad i \in \{x,y\} \tag{4.35}$$

follows that the partial derivatives $\partial\underline{l}_t/\partial\phi_{ij}$ can be expressed as

$$\partial\underline{l}_t/\partial\phi_{ij} = \left(\Gamma^{yi}\underline{\zeta}_{yt} + \Gamma^{xi}\underline{\zeta}_{xt}\right)\underline{v}_{jt}, \quad i,j \in \{x,y\}, \tag{4.36}$$

and the vector $\underline{m}_\phi$ is equal to

$$\begin{aligned} \underline{m}_\phi &= \sum\left(\partial\underline{l}_t/\partial\phi_{yy}', \partial\underline{l}_t/\partial\phi_{yx}', \partial\underline{l}_t/\partial\phi_{xy}', \partial\underline{l}_t/\partial\phi_{xx}'\right)' \\ &= \sum\left(\underline{\Gamma}^{-1}\underline{\zeta}_t \otimes \underline{V}_t\right). \end{aligned} \tag{4.37}$$

Each of the block matrices of the Hessian matrix

$$\underline{M} = \frac{1}{T}\sum \partial^2\underline{l}_t/\partial\vartheta\partial\vartheta' = \begin{pmatrix} \underline{M}_{\phi\phi} & \underline{M}_{\phi\Psi} \\ \underline{M}_{\Psi\phi} & \underline{M}_{\Psi\Psi} \end{pmatrix} \tag{4.38}$$

consists of 16 matrices. For example, the upper left block $\underline{M}_{\phi\phi}$ consists of the matrices

$$\underline{M}(\phi_{ij}, \phi_{hk}) = \frac{1}{T} \sum \partial^2 \underline{l}_t / \partial\phi_{ij} \partial\phi'_{hk}, \tag{4.39}$$

that may be estimated by

$$\underline{M}(\phi_{ij}, \phi_{hk}) = \underline{\Gamma}^{ih} \underline{v}_{jt} \underline{v}'_{kt}, \tag{4.40}$$

therefore

$$\underline{M}_{\phi\phi} = \frac{1}{T} \sum \left(\underline{\Gamma}^{-1} \otimes \underline{V}_t \underline{V}'_t \right). \tag{4.41}$$

Similarly,

$$\begin{aligned} \underline{M}_{\phi\Psi} &= \frac{1}{T} \sum \partial^2 \underline{l}_t / \partial\phi \partial\Psi' \doteq \frac{1}{T} \sum \left(\underline{\Gamma}^{-1} \underline{Z}'_t \otimes \underline{V}_t \right) \\ \underline{M}_{\Psi\phi} &= \underline{M}'_{\phi\Psi} \\ \underline{M}_{\Psi\Psi} &= \frac{1}{T} \sum \partial^2 \underline{l}_t / \partial\Psi \partial\Psi' \doteq \frac{1}{T} \sum \left(\underline{Z}_t \underline{\Gamma}^{-1} \underline{Z}'_t \right), \end{aligned} \tag{4.42}$$

where $Z_t = (z_{yt}, z_{xt})$ with $z_{it} = \partial F_i(X_t, \Psi_i)/\partial\Psi_i$, $i \in \{x, y\}$.

Parameter constancy test

Since the parameters of a smooth transition vector autoregressive model are estimated under the assumption of parameter constancy, it is important to develop the PC test in the multivariate setting. The null hypothesis of parameter constancy is tested against the alternative hypothesis of time-varying parameters of the form

$$\varphi_i(t) = \varphi_i + \lambda_{1i} H_i(t), \quad \theta_i(t) = \theta_i + \lambda_{2i} H_i(t), \quad i = x, y, \tag{4.43}$$

where

$$H_i(t) = (1 + \exp(-\gamma_{1i}(t^k + \nu_{i(k-1)} t^{k-1} + \ldots + \nu_{i1} t + \nu_{i0})))^{-1} - 0.5. \tag{4.44}$$

$H_i(t)$ thus enables nonmonotonic and asymmetric change in the parameters $\varphi_i(t)$ and $\theta_i(t)$. The transition function parameters γ_i and c_i are assumed constant also under the alternative hypothesis. Let us use the first order Taylor expansion around $\gamma_i = 0$ in place of $H_i(t)$ in equation (4.43). After substituting φ_i and θ_i for $\varphi_i(t)$ and $\theta_i(t)$, respectively, system (4.20) can be written as

$$\begin{aligned} y_t &= \omega'_{y0}A_{yt} + \omega'_{y1}A_{yt}t + \ldots + \omega'_{yk}A_{yt}t^k + \\ &+ \left(\tilde{\omega}'_{y0}\tilde{A}_{yt} + \tilde{\omega}'_{y1}\tilde{A}_{yt}t + \ldots + \tilde{\omega}'_{yk}\tilde{A}_{yt}t^k\right) G_y + v_{yt} \\ x_t &= \omega'_{x0}A_{xt} + \omega'_{x1}A_{xt}t + \ldots + \omega'_{xk}A_{xt}t^k + \\ &+ \left(\tilde{\omega}'_{x0}\tilde{A}_{xt} + \tilde{\omega}'_{x1}\tilde{A}_{xt}t + \ldots + \tilde{\omega}'_{xk}\tilde{A}_{xt}t^k\right) G_x + v_{xt} \end{aligned} \tag{4.45}$$

and the null hypothesis of constant parameters is of the form

$$H_0 : \omega_{i1} = \ldots = \omega_{ik} = 0,\ \tilde{\omega}_{i1} = \ldots = \tilde{\omega}_{ik} = 0,\ \ i = x, y. \tag{4.46}$$

Note that the vector A_{yt} contains only those elements of the vector X_t, whose coefficients in the linear part of the y_t equation are not assumed zero a priori, and the vector $\tilde{A}_{yt}$ in the nonlinear part of the same equation is defined in an analogous way. Having performed auxiliary regression (4.45), one can test for constant parameters with a standard Lagrange multiplier type test.

No remaining nonlinearity test

Following Eitrheim and Teräsvirta [18], Camacho introduces a smooth transition vector autoregressive model with 2 additive components

$$\begin{aligned} y_t &= \varphi'_y X_{yt} + (\theta'_y X_{yt})G^1_y(s^1_{yt}) + (\tilde{\theta}'_y X_t)G^2_y(s^2_{yt}) + u_{yt} \\ x_t &= \varphi'_x X_{xt} + (\theta'_x X_{xt})G^1_x(s^1_{xt}) + (\tilde{\theta}'_x X_t)G^2_x(s^2_{xt}) + u_{xt}. \end{aligned} \tag{4.47}$$

The test is performed in a similar way as the linearity test, i.e. the Taylor approximation of a suitable order is used instead of the transition function G^2_i. The auxiliary regression to be performed in case the transition variable $s^2_{xt} = s^2_{yt} = z_t$ belongs to X_t is

$$\begin{aligned} y_t &= \alpha'_y X_{yt} + \widetilde{\alpha}'_y X_{yt}G^1_y + \delta'_y \overline{X}_{yt} + \widetilde{\delta}'_y \overline{X}_{yt}G^1_y + \\ &+ \sum_{h=1}^{3} \xi'_{yh}\tilde{X}_t z^h_t + v_{yt} \\ x_t &= \alpha'_x X_{xt} + \widetilde{\alpha}'_x X_{xt}G^1_x + \delta'_x \overline{X}_{xt} + \widetilde{\delta}'_x \overline{X}_{xt}G^1_x + \\ &+ \sum_{h=1}^{3} \xi'_{xh}\tilde{X}_t z^h_t + v_{xt} \end{aligned} \tag{4.48}$$

and the null hypothesis of no remaining nonlinearity can be written as

$$H_0 : \xi_{i1} = \xi_{i2} = \xi_{i3} = 0,\ \delta_i = \widetilde{\delta}_i = 0,\ \ i = x, y. \tag{4.49}$$

New notation was introduced in the auxiliary regressions above. The vectors φ_i and θ_i are divided into two subvectors, namely α_i and $\widetilde{\alpha}_i$, respectively, with nonzero components, and δ_i and $\widetilde{\delta}_i$, respectively, that are assumed to be zero in the parameter estimation. This can be done without loss of generality. Analogously, the vector of explanatory variables is divided into subvectors X_{it} and $\overline{X}_{it}$. Obviously, the null hypothesis of no remaining nonlinearity can be tested with the Lagrange multiplier test.

Predictive accuracy of the estimated model

After the null hypothesis of linearity is rejected, one has to choose a suitable transition variable from a set of many available candidates. As already mentioned, Teräsvirta [73] suggests choosing the variable with the smallest p-value. However, Camacho points out two drawbacks of this heuristical method. Firstly, the decision is not clear when the p-values are similar. Secondly, even if the linearity is weekly rejected for one of the candidates for the transition variable, the obtained nonlinear model may have good forecasting properties. Camacho solves the problem by estimating the specified smooth transition vector autoregressive model for each of the transition variables rejecting the null of linearity and then choosing the best model on the basis of predictive accuracy. The type of the transition function for each of the transition variables is selected following the decision rule from section 3.2.2. Several measures of predictive accuracy are employed, namely

1. *Certain positive rate* (CPR) and *certain negative rate* (CNR)
   CPR (CNR) is defined as the percentage of quarters in which the model correctly predicts GDP rises (falls).

2. *False positive rate* (FPR) and *false negative rate* (FNR)
   FPR (FNR) is defined as the percentage of quarters in which the actual output growth is positive (negative) and the output growth as predicted by the model is negative (positive).

3. *Mean square error* (MSE)
   MSE is defined as

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{y}_t)^2$$

and measures the difference between the actual output growth y_t and the estimated output growth $\widehat{y_t}$.

4. *Turning points error* (TPE)

$$TPE = \frac{1}{T}\sum_{t=1}^{T}(d_t - \widehat{d_t})^2,$$

where d_t equals 1 in times of official recessions. Due to the fact that the value of the logistic transition function can be interpreted as the probability of expansion, the estimated probability of recession is given by $\widehat{d_t} = 1 - G_y$.

5. *Diebold-Mariano* (DM), *Modified Diebold-Mariano* (MDM), *Morgan-Granger-Newbold* (MGN) and *Meese-Rogoff* (MR) tests
MSE and TPE can be used as measures for comparing forecasting properties of several competing models. With the DM, MDM, MGN and MR tests one can test the null hypothesis of no difference in forecasting accuracy of these competing models.

4.3 Smooth transition vector autoregressive models with different transition variables

As already mentioned, Weise [80], van Dijk [78] and Camacho [10] all assume the same transition variable and the same type of the transition function in every equation of a smooth transition vector autoregressive model, with the interpretation that the economic practice imposes common nonlinear features. But this argument is not convincing, since such a conclusion cannot be derived from economic theory, while applied econometric studies analyzing nonlinear systems are scarce. For this reason we shall try to extend the work of Camacho by allowing different smooth transition functional forms in different equations.

When performing the system linearity test Camacho postulates a two-variable linear VAR(p) model under the null hypothesis and the smooth transition vector autoregressive model

$$\begin{aligned} y_t &= \varphi_y^{'} X_t + (\theta_y^{'} X_t) G_y(s_{yt}) + u_{yt} \\ x_t &= \varphi_x^{'} X_t + (\theta_x^{'} X_t) G_x(s_{xt}) + u_{xt}, \end{aligned} \tag{4.50}$$

where

$$X_t = (1, y_{t-1}, x_{t-1}, \ldots, y_{t-p}, x_{t-p})^{'} = (1, \tilde{X}_t^{'})^{'},$$

with the same transition variable $s_{yt} = s_{xt} = s_t$ and the same type of transition function $G_x = G_y = G$ under the alternative. After estimating the auxiliary regression

$$\begin{aligned} y_t &= X_t^{\prime}\eta_{y0} + (\tilde{X}_t^{\prime} s_t)\eta_{y1} + (\tilde{X}_t^{\prime} s_t^2)\eta_{y2} + (\tilde{X}_t^{\prime} s_t^3)\eta_{y3} + v_{yt} \\ x_t &= X_t^{\prime}\eta_{x0} + (\tilde{X}_t^{\prime} s_t)\eta_{x1} + (\tilde{X}_t^{\prime} s_t^2)\eta_{x2} + (\tilde{X}_t^{\prime} s_t^3)\eta_{x3} + v_{xt}, \end{aligned} \tag{4.51}$$

the null hypothesis

$$H_0 : \eta_{i1} = \eta_{i2} = \eta_{i3} = 0, \quad i = x, y, \tag{4.52}$$

(with the alternative of at least one of the coefficient vectors different from zero) is tested. The equations in (4.51) are estimated as a system with the method of maximum likelihood. Single equation estimators would also be consistent, although not efficient. The null hypothesis (4.52) can be tested with the LM test described in section 3.6.

If the restriction $s_{yt} = s_{xt} = s_t$ is not imposed, the system linearity test can be performed by testing the same null hypothesis, this time based on the auxiliary regression allowing for different transition variables

$$\begin{aligned} y_t &= X_t^{\prime}\eta_{y0} + (\tilde{X}_t^{\prime} s_{yt})\eta_{y1} + (\tilde{X}_t^{\prime} s_{yt}^2)\eta_{y2} + (\tilde{X}_t^{\prime} s_{yt}^3)\eta_{y3} + v_{yt} \\ x_t &= X_t^{\prime}\eta_{x0} + (\tilde{X}_t^{\prime} s_{xt})\eta_{x1} + (\tilde{X}_t^{\prime} s_{xt}^2)\eta_{x2} + (\tilde{X}_t^{\prime} s_{xt}^3)\eta_{x3} + v_{xt}. \end{aligned} \tag{4.53}$$

The system linearity test will be rejected if at least one of the relationships under observation is nonlinear, or more specifically, is characterized by smooth transition between parameter regimes. It is reasonable to believe that situations with only one of the equations being nonlinear can occur in the economic practice. Estimating both equations with the smooth transition specification would be inefficient in this case. To solve this problem, single equation linearity tests based on the system estimates of auxiliary regression (4.53) may be applied. For example, to develop the single equation linearity test for the first equation, the null hypothesis

$$H_0 : \eta_{y1} = \eta_{y2} = \eta_{y3} = 0 \tag{4.54}$$

should be verified. Testing such a null hypothesis corresponds to imposing the model

$$\begin{aligned} y_t &= \varphi_y^{'} X_t + u_{yt} \\ x_t &= \varphi_x^{'} X_t + (\theta_x^{'} X_t) G_x(s_{xt}) + u_{xt} \end{aligned} \tag{4.55}$$

under the null and model (4.50) with the STR specification in both equations under the alternative. The heuristic procedure for selecting the transition variable(s) can be derived similarly to the one explained by Teräsvirta [73] in the univariate setting:

1. Perform the system linearity test for each of the pairs of the possible transition variables in turn.

2. Carry out single equation linearity tests for each of the pairs of transition variables that reject the system linearity test.

3. **(i)** If there are pairs of transition variables for which the single equation tests reject the null hypothesis of linearity for both equations, choose the pair with the strongest rejection of the system linearity test.

   (ii) If for each of the pairs rejecting the null of system linearity only one of the single equation linearity tests is rejected, choose the pair of transition variables with the strongest rejection of the single equation test. Specify the corresponding equation as a smooth transition regression, while specifying the other equation as linear.

After the transition variables (or variable) have been chosen, the decision about the type of the transition function should be made. Camacho proposes a straightforward generalization of the sequence of nested hypotheses from section 3.2.2, namely

$$\begin{aligned} H_{04} &: \eta_{y3} = \eta_{x3} = 0, \\ H_{03} &: \eta_{y2} = \eta_{x2} = 0 |\ \eta_{y3} = \eta_{x3} = 0, \\ H_{02} &: \eta_{y1} = \eta_{x1} = 0 |\ \eta_{y2} = \eta_{x2} = \eta_{y3} = \eta_{x3} = 0, \end{aligned} \tag{4.56}$$

which can be tested with a sequence of F tests. The coefficient vectors η_{ij} refer to auxiliary regression (4.51). The decision rule determines the type of the transition function depending on the nested hypothesis with the strongest rejection (see section 3.2.2 for details). Note that the same functional form is selected for both equations.

If the assumption of the same type of the transition function in both equations, namely $G_x = G_y = G$, is also relaxed, the transition function is chosen for each equation separately. In case of the first equation, the regressions in (4.53) are estimated with a system estimation method and the following sequence of nested hypotheses is verified:

$$\begin{aligned} H_{04} &: \eta_{y3} = 0, \\ H_{03} &: \eta_{y2} = 0 |\, \eta_{y3} = 0, \\ H_{02} &: \eta_{y1} = 0 |\, \eta_{y2} = \eta_{y3} = 0. \end{aligned} \tag{4.57}$$

When the second equation has a linear specification, auxiliary regression (4.53) should be modified accordingly, with the coefficient vectors η_{x1}, η_{x2} and η_{x3} set to zero prior to estimation.

An alternative specification procedure would follow the suggestion of Camacho and specify a model for each pair of the transition variables rejecting the null hypothesis of system linearity. The final model can be selected on the basis of forecasting power or any other measure of the model adequacy. Of course this strategy is more time consuming, especially if the null hypothesis of system linearity is rejected for several pairs of transition variables. Our approach extends the modeling cycle developed by Camacho, since the set of models considered for smooth transition specification includes all of the models studied by Camacho.

To illustrate the proposed heuristic procedure, we shall apply it to the data analyzed by Camacho and obtained from his web site (see appendix C). The letters y_t and x_t are used for the two variables in the autoregressive system, since the smooth transition approach is applied to the rate of growth of US GDP and the rate of growth of US composite index of leading indicators, respectively. Both series include quarterly observations from 1959:1 to 2002:1.

Let us first derive the final model of Camacho. Using the information criteria, a linear VAR(1) model is specified and subjected to the the system linearity tests. In addition to y_{t-1} and x_{t-1}, the variables y_{t-2} and x_{t-2} are also regarded as candidates for the transition variable. The results are given in the corresponding rows of Table 4.1. As the null of system linearity is rejected in all four cases, Camacho estimates four smooth transition autoregressive models. The type of the transition function for each of the transition variables is selected using the decision rule explained on page 74 and is the same for both equations. The final model is decided upon on the basis of predictive accuracy (see previous section for details). The maximum likelihood estimates of the final model given below are slightly different from those in the paper by Camacho [10]:

$$\begin{aligned} y_t &= \underset{(0.092)}{0.603} \cdot x_{t-1} + \underset{(0.200)}{0.900} * \left(\frac{1}{1+e^{-\underset{(1.359)}{1.840}(y_{t-2}-\underset{(0.393)}{0.116})}} \right) \\ x_t &= \underset{(0.126)}{0.415} - \underset{(0.154)}{0.364} \cdot y_{t-1} + \underset{(0.060)}{0.507} \cdot x_{t-1} + \\ &+ \left(-\underset{(0.252)}{0.308} + \underset{(0.213)}{0.324} \cdot y_{t-1}\right) * \left(\frac{1}{1+e^{-\underset{(10.100)}{4.413}(y_{t-2}-\underset{(0.343)}{0.373})}} \right). \end{aligned} \tag{4.58}$$

As pointed out by Teräsvirta, precise joint estimation of the slope parameter and the threshold can be problematic.

When the restrictions regarding the transition variable and the type of the transition function are omitted, the linearity tests have to be carried out for every pair of the possible transition variables. The results of the system linearity tests as well as single equation tests are given in Table 4.1.

It can be observed that the single equation linearity test is strongly rejected for every pair of the transition variables when testing the second equation, while this holds true only for the pair x_{t-2}, y_{t-1} in case of the first equation. Thus the rejection of system linearity is due mainly to the nonlinear features in the second relation. Following the previously explained heuristic procedure, we choose the transition variables x_{t-2} and y_{t-1} for the first and second equation, respectively.

Next, the question of the type of the transition function in each of the equations has to be investigated. The sequence of nested hypotheses (4.57) is performed with the results given in Table 4.2. The F2 test yields the lowest p-value for both equations thus indicating the LSTR1 transition model in both cases.

Tvar in 1. and 2. eq.	LINEARITY TESTS (p-values) System test	First eq. test	Second eq. test
x_{t-1}, x_{t-1}	0.0020	0.4137	0.0017
x_{t-1}, x_{t-2}	0.0091	0.1882	0.0103
x_{t-1}, y_{t-1}	0.0039	0.1475	0.0038
x_{t-1}, y_{t-2}	0.0076	0.1332	0.0083
x_{t-2}, x_{t-1}	0.0003	0.0774	0.0004
x_{t-2}, x_{t-2}	0.0037	0.0755	0.0082
x_{t-2}, y_{t-1}	0.0008	0.0300	0.0014
x_{t-2}, y_{t-2}	0.0033	0.0565	0.0073
y_{t-1}, x_{t-1}	0.0041	0.6775	0.0003
y_{t-1}, x_{t-2}	0.0553	0.8199	0.0075
y_{t-1}, y_{t-1}	0.0315	0.8003	0.0035
y_{t-1}, y_{t-2}	0.0529	0.7331	0.0070
y_{t-2}, x_{t-1}	0.0018	0.3793	0.0005
y_{t-2}, x_{t-2}	0.0119	0.2449	0.0054
y_{t-2}, y_{t-1}	0.0057	0.2106	0.0022
y_{t-2}, y_{t-2}	0.0114	0.2005	0.0052

Table 4.1: Linearity test results (p-values)

	NESTED TESTS (p-values)					
Tvar in 1.	First equation			Second equation		
and 2. eq.	F4	F3	F2	F4	F3	F2
x_{t-2}, y_{t-1}	0.2250	0.3676	0.0112	0.6859	0.0370	0.0008

Table 4.2: Tests for choosing the type of transition function (p-values)

The maximum likelihood estimates of the specified smooth transition vector autoregressive model are given by

$$
\begin{aligned}
y_t &= \underset{(0.103)}{0.467} \cdot x_{t-1} + \underset{(0.371)}{1.213} * \left(\frac{1}{1 + e^{-\underset{(0.900)}{1.389}(x_{t-2} - \underset{(0.523)}{0.053})}} \right) \\
x_t &= -\underset{(2.891)}{1.561} + \underset{(0.059)}{0.475} \cdot x_{t-1} - \underset{(2.119)}{1.680} \cdot y_{t-1} + \\
&+ \underset{(1.580)}{1.489} \cdot y_{t-1} * \left(\frac{1}{1 + e^{-\underset{(3.288)}{2.592}(y_{t-1} + \underset{(0.994)}{0.614})}} \right)
\end{aligned}
\tag{4.59}
$$

It can be deducted from Table 4.3 that models (4.58) and (4.59) are comparable in terms of fit, if model (4.59) is not even slightly better. Therefore it would be wise to consider also (4.59) when searching for the model with the best forecasting properties.

	Model	First equation		Second equation	
Value	*logL*	R^2	*S.E.*	R^2	*S.E.*
Model (4.58)	-327.14	0.352	0.754	0.253	0.718
Model (4.59)	-322.20	0.369	0.744	0.287	0.701

Table 4.3: Comparing the fit of models (4.58) and (4.59)

The modeling procedure proposed by Camacho has several drawbacks. Firstly, the system linearity test based on auxiliary regression (4.51) is rejected if at least one of the equations includes nonlinear terms. Specifying every equation as nonlinear based only on the rejection of the system linearity test thus neglects the possibility of a system involving linear and nonlinear equations and can yield inefficient estimates. In chapter 6 we develop a nonlinear monetary model of inflation characterized by the real money demand equation augmented by the Phillips curve and the equation of Okun's law. It turns out that while the money demand equation and the Phillips curve exhibit nonlinear features, the Okun's law should be specified as a linear equation, since the single equation linearity tests cannot be rejected in any case. Additionally, the regime changes in the money demand equation and in the Phillips curve are governed by different transition variables.

Secondly, the limitations in the specified functional form can be justified neither by the relationships postulated within the economic theory nor by taking into account the very few existing applied studies. In chapter 5 we present an application of the common nonlinearities approach to analyze the components of the real exchange rates of the Slovenian Tolar versus the currencies of its five major trading partners. It turns out that in only one of the five cases examined one common nonlinear component can adequately describe the development of the three real exchange rate components, while in other four cases a model involving at least two nonlinear terms should be specified. Thus, the specification with the same transition variable and the same type of the transition function in every equation of a smooth transition vector autoregressive model is too restrictive and should not be imposed a priory.

Chapter 5

Real Tolar - Euro exchange rate model

In this chapter, three-variable smooth transition vector autoregressive models of the consumer price index for Slovenia, consumer price index of another country and the nominal exchange rate between the currencies of both countries are discussed. The investigation applies the common nonlinearities techniques to small models of the real exchange rate, decomposed into its three components, domestic prices (P_t), foreign prices (P_t^*) and the nominal exchange rate (S_t). The empirical investigation includes five most important foreign trade partners of Slovenia, namely Germany, Italy, France, Austria and Croatia.

Monthly data for the period from January 1988 till December 2003 were obtained from the Bank of Slovenia and from the Statistical Office of the Republic of Slovenia. Due to the fact that Slovenia declared independence in June 1991 and introduced its own currency (Tolar) in October of the same year, only the data for the period from January 1993, when Tolar was already an established currency, were used in the study. In the case of Croatia, the period under investigation was shortened additionally because of the war. Only the data from April 1995 till December 2003 were taken into account. The econometric model employs variables expressed in growth rates with the help of the logarithmic transformation, therefore small letters are used to denote the transformed variables.

In a preliminary specification, all equations were modeled as linear relationships. This simplifies the search for an appropriate nonlinear specification. Firstly, unit root tests were applied to the variables p_t, p_t^* and s_t

for each of the countries. All of the variables turned out to be integrated of order 1, or I(1), as one cannot reject the null hypothesis of unit root for any variable. Next, cointegration tests were performed and the linear vector error correction models (VECM) were specified. The null hypothesis of no cointegrating relations could not be rejected only in case of Croatia, therefore a linear VAR model in the differenced variables Δp_t, Δp_t^* and Δs_t was specified. Orthogonal seasonal dummy variables, denoted by d1 to d12, were introduced into some of the models to reduce the autocorrelation effects. Test results of the preliminary linear specification and the estimates of the VECM models are given in appendix B.

5.1 Linearity test results

In order to improve specification we investigate the influence of nonlinearities, which we assume to be of the smooth transition kind. For this purpose, we test the null hypothesis of linearity against the alternative of a smooth transition autoregressive model for each of the equations and each of the possible transition variables in turn. The values of the F-statistic (and the corresponding p-values in brackets) are given in Table 5.1 below. The cointegrating equations are denoted by ce1 and ce2.

Germany			
transition variable	Δp equation	Δp^* equation	Δs equation
$ce1_t$	1.642 (0.039)	1.072 (0.393)	4.178 (0.000)
Δp_{t-1}	1.135 (0.321)	1.251 (0.211)	1.819 (0.017)
Δp^*_{t-1}	0.847 (0.699)	0.567 (0.965)	1.383 (0.124)
Δs_{t-1}	1.397 (0.117)	0.995 (0.492)	2.612 (0.000)
Δp_{t-2}	1.936 (0.009)	0.554 (0.971)	2.566 (0.000)
Δp^*_{t-2}	1.034 (0.440)	1.432 (0.101)	0.791 (0.772)
Δs_{t-2}	1.601 (0.047)	0.684 (0.889)	4.175 (0.000)
Δp_{t-3}	0.711 (0.864)	1.022 (0.457)	2.037 (0.006)
Δp^*_{t-3}	1.387 (0.122)	1.264 (0.200)	1.352 (0.141)
Δs_{t-3}	1.146 (0.309)	0.456 (0.993)	3.471 (0.000)

Italy			
transition variable	Δp equation	Δp^* equation	Δs equation
$ce1_t$	1.480 (0.125)	0.898 (0.574)	3.104 (0.000)
Δp_{t-1}	1.123 (0.347)	1.558 (0.098)	1.494 (0.120)
Δp^*_{t-1}	2.797 (0.001)	2.757 (0.001)	1.809 (0.042)
Δs_{t-1}	1.681 (0.065)	1.022 (0.442)	2.248 (0.009)

France			
transition variable	Δp equation	Δp^* equation	Δs equation
$ce1_t$	1.213 (0.250)	0.645 (0.917)	3.172 (0.000)
Δp_{t-1}	0.878 (0.655)	0.935 (0.576)	1.334 (0.160)
Δp^*_{t-1}	0.665 (0.901)	1.3710 (0.139)	0.642 (0.919)
Δs_{t-1}	1.210 (0.253)	0.790 (0.771)	2.527 (0.001)
Δp_{t-2}	1.276 (0.199)	0.760 (0.806)	1.160 (0.300)
Δp^*_{t-2}	1.022 (0.460)	0.829 (0.721)	0.682 (0.887)
Δs_{t-2}	1.857 (0.017)	0.939 (0.571)	3.609 (0.000)
Δp_{t-3}	1.112 (0.351)	0.776 (0.787)	2.226 (0.003)
Δp^*_{t-3}	0.861 (0.677)	0.973 (0.524)	1.249 (0.220)
Δs_{t-3}	1.150 (0.310)	0.843 (0.701)	2.045 (0.007)

Austria			
transition variable	Δp equation	Δp^* equation	Δs equation
$ce1_t$	1.392 (0.153)	1.002 (0.468)	2.752 (0.001)
$ce2_t$	2.619 (0.001)	1.712 (0.043)	3.423 (0.000)
Δp_{t-1}	1.284 (0.216)	1.207 (0.272)	1.817 (0.034)
Δp^*_{t-1}	1.356 (0.172)	1.146 (0.323)	1.655 (0.061)
Δs_{t-1}	1.765 (0.041)	1.355 (0.173)	4.024 (0.000)

Croatia			
transition variable	Δp equation	Δp^* equation	Δs equation
Δp_{t-1}	1.222 (0.282)	1.432 (0.166)	0.840 (0.617)
Δp^*_{t-1}	0.755 (0.704)	1.110 (0.365)	2.719 (0.004)
Δs_{t-1}	0.616 (0.834)	0.638 (0.814)	4.208 (0.000)

Table 5.1: F-values (and p-values) for testing linearity against STR

The goal of the study is to obtain a model with only one common nonlinear component, because in this way a parsimonious specification is achieved.

Obviously, the necessary condition for the existence of such a model is a transition variable for which the null hypothesis of linearity is rejected for every equation in the model. There are only two variables that comply with this condition in our case, namely the variable Δp^*_{t-1} in the model for Italy and the variable $ce2_t$ in the model for Austria (see Table 5.1). The significance level of 5 % is assumed. Table 5.2 shows the results of the common nonlinearities test for both of the mentioned transition variables.

Italy		
tvar: Δp^*_{t-1}		
s	p-value	df
1	0.230	10
2	0.033	22
3	0.000	36

Austria		
tvar: $ce2_t$		
s	p-value	df
1	0.262	13
2	0.079	28
3	0.001	45

Table 5.2: Common nonlinearities test for Italy and Austria

In accordance with the theory from section 4.1, s equals 1 for Italy and 2 for Austria. The number of common nonlinear components, which is determined by the formula n-s, takes the value of 2 for Italy and the value of 1 for Austria. Consequently, we shall concentrate on the model for Austria from now on.

5.2 Estimated model for Austria

The full information maximum likelihood (FIML) estimator was employed to allow for correlated residuals in different equations. Due to occasional problems with convergence of the nonlinear optimization procedure some experimentation to find appropriate starting values was required. The final

set of estimates obtained for the parameters of the nonlinear logistic smooth transition vector error correction model of the Slovenian Tolar versus the Austrian Schilling, respectively Euro, can be found below. The estimated standard errors are given in brackets.
Common nonlinear component:

$$
\begin{aligned}
com_t &= \frac{1}{1+e^{-\underset{(29.608)}{53.0123}(ce2_t+\underset{(0.016)}{0.0993})}} * \\
&* \Big(\underset{(0.0096)}{-0.0315} - \underset{(0.0196)}{0.0717} \cdot ce1_t - \underset{(0.0613)}{0.2177} \cdot ce2_t + \\
&+ \underset{(0.2486)}{0.4792} \cdot \Delta p^*_{t-1} + \underset{(0.0573)}{0.1063} \cdot \Delta s_{t-1}\Big)
\end{aligned}
\tag{5.1}
$$

First equation:

$$
\begin{aligned}
\Delta p_t &= \underset{(0.0091)}{0.0384} + \underset{(0.0188)}{0.0388} \cdot ce1_t + \underset{(0.0663)}{0.1506} \cdot ce2_t - \underset{(0.2491)}{0.6346} \cdot \Delta p^*_{t-1} + \\
&+ \underset{(0.0015)}{0.0086} \cdot d1_t + \underset{(0.0020)}{0.0055} \cdot d2_t + \underset{(0.0018)}{0.0053} * d3_t + \underset{(0.0015)}{0.0053} \cdot d4_t + \\
&+ \underset{(0.0015)}{0.0056} \cdot d5_t + \underset{(0.0019)}{0.0033} \cdot d7_t + \underset{(0.0016)}{0.0057} \cdot d9_t + \underset{(0.0025)}{0.0036} \cdot d10_t + \\
&+ \underset{(0.0015)}{0.0047} \cdot d11_t + \underset{(0.0018)}{0.0025} \cdot d12_t + com_t
\end{aligned}
\tag{5.2}
$$

Second equation:

$$
\begin{aligned}
\Delta p^*_t &= \underset{(0.0010)}{0.0031} + \underset{(0.0052)}{0.0072} \cdot ce2_t + \underset{(0.0390)}{0.0526} \cdot \Delta s_{t-1} + \underset{(0.0013)}{0.0016} \cdot d2_t + \\
&+ \underset{(0.0009)}{0.0013} \cdot d7_t - \underset{(0.0009)}{0.0027} \cdot d9_t + \underset{(0.0147)}{0.0620} \cdot com_t
\end{aligned}
\tag{5.3}
$$

Third equation:

$$\begin{aligned}
\Delta s_t &= \underset{(0.0147)}{0.0620} + \underset{(0.0305)}{0.1642} \cdot ce1_t + \underset{(0.1055)}{0.4629} \cdot ce2_t - \underset{(0.4513)}{1.1531} \cdot \Delta p^*_{t-1} + \\
&+ \underset{(0.0905)}{0.5001} \cdot \Delta s_{t-1} - \underset{(0.0017)}{0.0024} \cdot d5_t - \underset{(0.0020)}{0.0022} \cdot d7_t + \underset{(0.6469)}{1.9444} \cdot com_t
\end{aligned} \tag{5.4}$$

Note that both of the crucial parameters in the common nonlinear part, γ and c, are significant at the 10 % level. The γ value of approximately 53 indicates rapid transition between the two extreme regimes.

A comparison of single equations from the linear and nonlinear system (Table 5.3) reveals an increase in explanatory power for equations 1 and 3 (R^2 increases from 0.62 to 0.65 and from 0.67 to 0.76, respectively) and a decrease in the standard error of regression from 0.0039 to 0.0038 and from 0.0051 to 0.0043, respectively. For equation 2, the situation is just the opposite. There is a decrease in explanatory power, while the standard error of regression stays the same. It should be emphasized that there are better ways to analyze systems of equations than single-equation comparison. If we compare the value of the log likelihood for both systems, we can observe that it is higher for the nonlinear system (1668.52 as compared to 1649.64 for the linear system), indicating an improvement in specification.

Linear system			
	Δp equation	Δp^* equation	Δs equation
R^2	0.62	0.22	0.67
S.E.	0.0039	0.0028	0.0051

Nonlinear system			
	Δp equation	Δp^* equation	Δs equation
R^2	0.65	0.17	0.76
S.E.	0.0038	0.0028	0.0043

Table 5.3: Comparing equations in the linear and nonlinear system

5.3 Conclusion

The obtained real exchange rate model of the Slovenian Tolar versus Austrian Schilling, respectively Euro, contains only one common nonlinear component, as desired. On the other hand, models for Germany, Italy, France and Croatia cannot be adequately described with the help of only one type of nonlinearity. One of the possible explanations for such results could be the late accession of Austria to the European Union. Austria joined EU in the year 1995, whereas Germany, Italy and France were already member states in the year 1993, when we started our investigation. The process of Austria's EU accession had a deep impact on its economic structure and the relation to its neighbour states. In particular prices have been severely affected. These adjustments together with those ongoing in neighbouring Slovenia seem to be captured by a common nonlinear factor in the components of the real exchange rate. It concerns especially the effects of the lagged Austrian inflation rate and the nominal exchange rate, besides the cointegration terms. This lends economic support to the specification of a logistic smooth transition model with only one common nonlinear component.

Chapter 6

A nonlinear monetary model of inflation

6.1 Theoretical specification

We start with a model of monetary inflation theory which can be shortly characterized by an equation describing the monetary system augmented by a Phillips curve and the equation of Okun's law. The simple elementary system is given by (see [23]):

$$\begin{aligned} m_t &= x_t + \pi_t, \\ \pi_t &= \pi_t^* - b(u_t - u^*), \\ u_t - u_{t-1} &= -a(x_t - x_t^*). \end{aligned} \tag{6.1}$$

These three equations determine three unknown variables: real growth rate (x_t), inflation rate (π_t) and unemployment rate (u_t) in terms of given monetary growth. The first equation representing the quantity equation in growth rates (assuming constant velocity) will be substituted by a demand for real money equation, which may not be homogenous of degree 0 in its arguments. We assume the nominal money stock to be given by the monetary authority. The equilibrium in the monetary sector can be described by a general equation expressing the relationship between money stock, output and prices, respectively interest rates. The Phillips curve relates inflation to the deviation of the unemployment rate from its natural rate (u^*) augmented by backward and forward inflationary expectations. Okun's law provides a relationship between the change in unemployment rate and the deviation of the actual from the trend rate of real output growth (x_t^*). By considering an Okun type relationship and additional supply effects in the Phillips curve the resulting model exhibits essential features of the "triangle model" as defined by Gordon [26]. Supply effects may include tax changes, rates of change in import or oil prices, or the change in the unemployment rate which may give rise to hysteresis. The (excess) demand side is typically represented by the unemployment gap and may also incorporate lagged effects of the growth of the stock of money.

This monetary approach to an explanation of inflation will be applied to seasonally adjusted quarterly data from West Germany between 1970:1 and 1998:4.

6.1.1 About the money demand and the quantity equation

Let us denote the stock of money by M. Money is understood by its narrow definition as the currency in circulation plus the sight deposits. This sum is usually denoted by M_1.

As explained in Burda and Wyplosz [8], the demand for nominal money is proportional to the price level (P), or, in other words, the demand for money is the demand for real money. The money demand also depends on the real output (X) and the nominal interest rate (r). Obviously, the higher the real output, the higher the money demand and the higher the interest rate, the lower the money demand. The relationship can be written in a compact form as

$$\frac{M}{P} = f(X, r). \tag{6.2}$$

The velocity of money (V) is a measure of how many times (on average) a unit of money is spent during a fixed time period, which is usually a year. V is defined by the so-called quantity equation as the ratio

$$V = \frac{P \cdot X}{M}. \tag{6.3}$$

If, for example, $V = 2$, then we say that the stock of money turns around twice a year. In the quantity theory the velocity of money is assumed constant, since the price level is determined by the quantity of money. Expressing equation (6.3) in terms of M,

$$M = \frac{P \cdot X}{V}, \tag{6.4}$$

and taking logarithms under the assumption of constant velocity (with the value of 1) yields

$$m = \pi + x. \tag{6.5}$$

Small letters were used to denote the growth rates in the logarithmic form. Thus, the first equation of system (6.1) is the quantity equation. For the purpose of estimation it will be substituted by the demand for money equation and referred to as the "surrogate quantity equation".

6.1.2 About the Phillips curve

In 1958, the British economist A.W. Phillips [62] published his famous paper about the relationship between the change of money wages and the rate of unemployment. Based on the UK data for the period from 1861 till 1957, he concluded that the dependency between the observed variables can be described by a downward sloping curve. Two years later, Samuelson and Sollow [64] modified the concept of the Phillips curve to represent a relationship between the rate of inflation and the rate of unemployment. Since the relationship was supposed to be stable, the Phillips curve became increasingly popular as a policy instrument allowing the policy makers to choose between alternative combinations of inflation and unemployment. The implied trade-off between inflation and unemployment meant that a reasonably high rate of inflation could be tolerated as this would lead to lower unemployment.

However, in the 1970's, several countries experienced stagflation - a phenomenon of a high inflation combined with a high unemployment rate. Obviously, this was in contradiction with the theory of a stable Phillips curve. As a consequence, new theories emerged, explaining how stagflation could occur. Milton Fridman proposed the so-called expectations-augmented Phillips curve. The equation of the Phillips curve was expanded to include a parameter for the expected rate of inflation (π_t^*),

$$\pi_t = \pi_t^* - b(u_t - u^*), \tag{6.6}$$

implying that different expected rates of inflation correspond to alternative Phillips curves. Linear Phillips curve (6.6) can be written more generally in the form

$$\pi_t = \pi_t^* + f(u_t). \tag{6.7}$$

As depicted in Figure 6.1, a change in the expected rate of inflation shifts the Phillips curve. The augmented theory distinguishes between the short-term

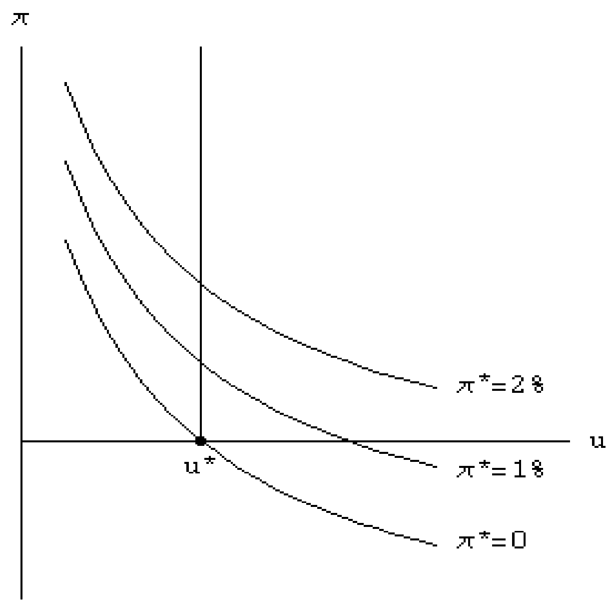

Figure 6.1: Expectations augmented Phillips curve

and the long-term Phillips curve. In the short run given by equation (6.6) there is a trade-off between the inflation rate and the unemployment rate, while in the long run the Phillips curve can be depicted by a vertical line (Figure 6.1). To establish this claim, one only has to observe that in the long run expectations are fully adjusted and there is no deviation of the actual rate of inflation from the expected rate of inflation. Thus, the vertical Phillips curve given by the formula $u_t = u^*$ can easily be derived from equation (6.6).

6.1.3 About the Okun‘s law

Okun's law describes the short-run relationship between the GDP gap and the unemployment rate. This empirical relationship, developed by A.M. Okun in the 1970‘s, can be stated as follows (compare Frisch [23]):

$$u_t = u^* - a\left(\frac{X_t - X_t^*}{X_t^*}\right), \tag{6.8}$$

where $a > 0$ is a constant term, u_t and u^* denote the actual and the natural rate of unemployment, X_t stands for the actual real output and X_t^* for the potential real output. Okun's estimate of the parameter a for the United States was approximately 0.3, implying that for each percentage point by which the unemployment rate is higher than the natural rate, real output is approximately 3.3 % lower than the potential real output. As a slightly more general relationship between the rate of unemployment and the rate of growth of real output we can write

$$u_t = u_{t-1} - a\Big(x_t - x_t^*\Big), \tag{6.9}$$

with x_t^* denoting the expected rate of real growth following the long-run trend. The difference $x_t - x_t^*$ is usually called the output gap.

6.2 Linear econometric model

The econometric model employs transformations of variables into growth rates. In a preliminary specification, all equations are modeled as linear relationships. This simplifies the search for an appropriate nonlinear specification. The starting point of our empirical investigation is the system

$$\begin{aligned} m_t &= x_t + \pi_t, \\ \pi_t &= \pi_t^* - b(u_t - u^*), \\ u_t &= u_{t-1} - a(x_t - x_t^*), \end{aligned} \tag{6.10}$$

which we shall augment with additional explanatory variables and adjustment terms, thus allowing the dynamic specification. On the basis of these empirical results further investigations will have to reveal any remaining nonlinearity in these relations.

6.2.1 Surrogate quantity equation

The first equation specifies real money demand growth (m_t) as dependent variable explained by adjustment terms representing the adjustment of real money growth to output growth (x_t) and prices. The dynamic linear specification of the surrogate quantity equation is of the form

$$m_t = a + b(L)x_t + c(L)\Delta\pi_{t-1} + d(L)m_{t-1} + f(L)\pi_t^f + \varepsilon_t. \tag{6.11}$$

The forward looking price expectation (π_t^f) is defined on page 96. A maximum of 4 lags was allowed in the lag polynomials $b(L)$, $c(L)$, $d(L)$ and $f(L)$,

SURROGATE QUANTITY EQUATION				
Dependent Variable: m_t				
Sample (adjusted): 1970:1 1998:4 (Observations: 116)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
$const.$	0.023727	0.005669	4.185161	0.0001
x_t	0.183880	0.127073	1.447037	0.1508
x_{t-4}	0.211157	0.098361	2.146751	0.0341
$\Delta\pi_{t-1}$	-0.582542	0.264546	-2.202045	0.0298
m_{t-1}	0.883156	0.047842	18.45990	0.0000
m_{t-4}	-0.267305	0.050737	-5.268417	0.0000
π_t^f	-0.005449	0.001437	-3.792490	0.0002
$dummy1$	0.112028	0.016734	6.694747	0.0000
R^2 = 0.855, S.E. = 0.022, SSR = 0.054, AIC = -4.690				

Table 6.1: OLS results for surrogate quantity equation (6.11)

thus extending over a period of a year because of the quarterly data. The least squares estimates (with the insignificant variables removed) are given in Table 6.1.

Results of the LM test of no error autocorrelation in Table 6.2 do not indicate autocorrelation, nor is there any evidence of ARCH effects. The obtained linear equation also proved satisfactory after being tested for normality of errors, autocorrelations, heteroscedasticity and misspecification, but the CUSUM and CUSUMQ tests indicate problems with parameter constancy. The CUSUM and CUSUMQ tests are based on the cumulative sum of the recursive residuals and the cumulative sum of the squared recursive residuals, respectively. The tests reject the null hypothesis of parameter constancy when the value of the cumulative sum statistic lies outside the area between the two critical lines establishing the 5 % significance level. The plots in Figure 6.2 and Figure 6.3 both reveal values of the test statistic outside the critical lines in the period from 1991:3 till 1993:1 and thus suggest structural breaks in the surrogate quantity equation.

Test	Jarque-Bera	White (cross terms)	Ljung-Box (4 lags)
Test statistic (p-value)	1.0705 (0.5855)	32.8128 (0.2854)	3.5264 (0.4740)
Test	RESET (1 fitted term)	RESET (2 fitted terms)	RESET (3 fitted terms)
Test statistic (p-value)	2.6805 (0.1016)	2.8020 (0.2464)	3.2566 (0.3537)
Test	Serial LM (1 lag)	Serial LM (2 lags)	Serial LM (4 lags)
Test statistic (p-value)	2.4817 (0.1152)	2.5010 (0.2864)	4.1963 (0.3862)
Test	ARCH LM (1 lag)	ARCH LM (2 lags)	ARCH LM (4 lags)
Test statistic (p-value)	0.1385 (0.7098)	0.2339 (0.8896)	1.5566 (0.6693)

Table 6.2: Specification and diagnostic tests for linear surrogate quantity equation (6.11)

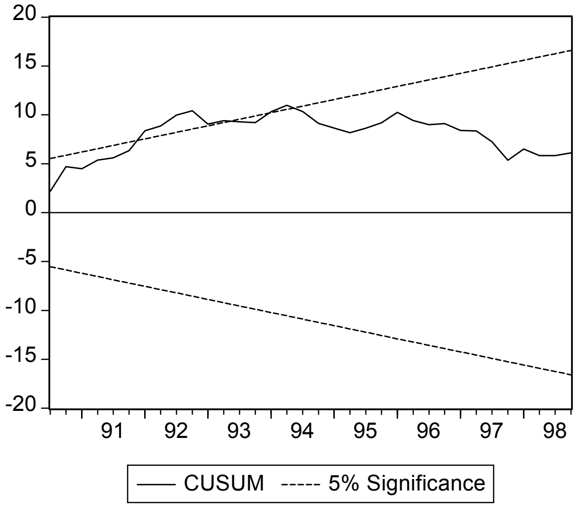

Figure 6.2: CUSUM test for linear surrogate quantity equation (6.11)

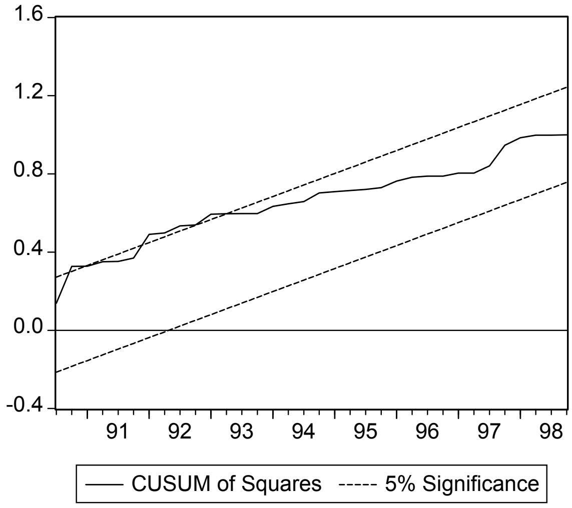

Figure 6.3: CUSUMQ test for linear surrogate quantity equation (6.11)

Figure 6.4 depicts recursive coefficient estimates of the linear surrogate quantity equation. Recursive estimates of a parameter vector are obtained by least squares estimation over gradually increasing periods. Sudden changes in the course of the recursive estimates imply structural change, whereas smooth changes hint at misspecification. The two-standard-error bands around each recursive coefficient are also shown in the plot. In our case, the coefficients $C(1)$ and $C(7)$ of the constant and the π_t^f variable display the most variation. The surrogate quantity equation is a potential candidate for nonlinear STR specification, since several coefficients do not seem to be constant over time.

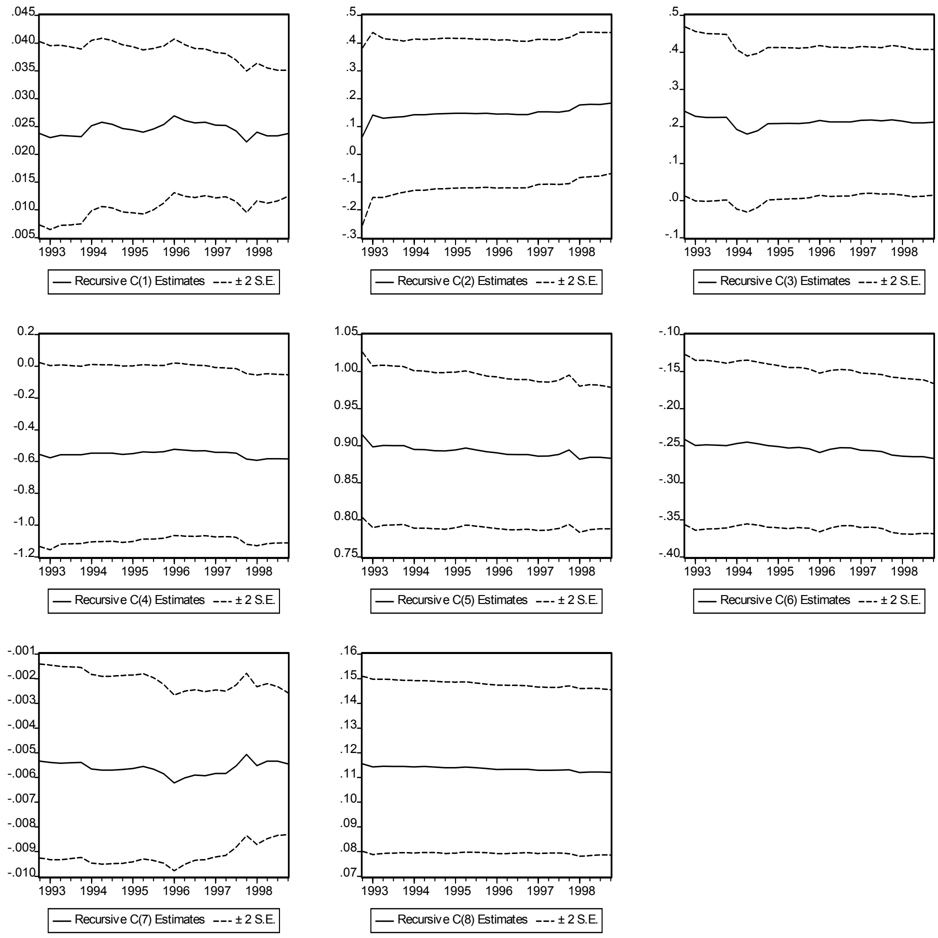

Figure 6.4: Recursive coefficients for linear surrogate quantity equation (6.11) (following the estimated coefficients in Table 6.1 row-wise)

6.2.2 Phillips curve

PHILLIPS CURVE				
Dependent Variable: $\Delta\pi_t$				
Sample (adjusted): 1969:3 1998:4 (Observations: 118)				
Variable	Coefficient	St. Error	t-Statistic	Prob.
$const.$	0.007590	0.002537	2.991486	0.0035
$\Delta\pi_{t-1}$	0.418053	0.087216	4.793325	0.0000
π_{t-1}	-0.247808	0.055656	-4.452521	0.0000
π_{t-4}	-0.166698	0.067031	-2.486884	0.0145
$\Delta\pi_{t-8}$	-0.089575	0.056247	-1.592533	0.1144
$\Delta\pi_{t-12}$	-0.110270	0.055852	-1.974331	0.0511
π_{t-5}	0.165211	0.053915	3.064292	0.0028
π_t^f	0.000746	0.000404	1.846239	0.0678
π_t^0	0.176137	0.014223	12.38415	0.0000
π_{t-1}^0	-0.203229	0.024364	-8.341309	0.0000
π_{t-2}^0	0.098185	0.019734	4.975404	0.0000
λ_{t-2}	-0.073331	0.026445	-2.772931	0.0066
u_t	-0.000693	0.000218	-3.174645	0.0020
m_{t-2}	0.029236	0.015522	1.883461	0.0625
m_{t-3}	-0.074374	0.021147	-3.516986	0.0007
m_{t-4}	0.064679	0.016840	3.840904	0.0002
$dummy2$	0.010957	0.002014	5.440388	0.0000
R^2 = 0.809, S.E. = 0.004, SSR = 0.002, AIC = -8.027				

Table 6.3: OLS results for linear Phillips curve (6.13)

The Phillips curve is modeled according to considerations in Böhm [6]. The inflation rate (π_t) depends on the unemployment gap ($u_t - u^*$), energy price inflation (π_t^0) and expected inflation modeled by backward and forward looking components. Forward looking price expectation (π_t^f) equals the difference between the nominal and the real rate of interest according to Fisher's formula

$$r_t^n = r_t^f + \pi_t^f. \tag{6.12}$$

As nominal rate, we use the long term government bond yield while the real interest rate is represented by a proxy variable, the ratio of the sum of real GDP over four quarters to the corresponding sum of real investment expenditures. In view of the modification of the Phillips curve by Samuelson and Solow [64], labour productivity growth (λ_t) is also considered in the equation. Thus, the Phillips curve in its linear specification including several lagged effects is given by

$$\begin{aligned} \pi_t &= a + b(L)\pi_{t-1} + c(L)\Delta\pi_{t-1} + d(L)\pi_t^f + f(u_t - u^*) + \\ &+ g(L)\pi_t^0 + h(L)\lambda_t + j(L)m_{t-1} + \varepsilon_t, \end{aligned} \tag{6.13}$$

where the degrees of the lag polynomials $b(L)$ and $c(L)$ were a priori limited by the value of 12 and the degrees of the polynomials $d(L)$, $g(L)$, $h(L)$ and $j(L)$ by the value of 4. The estimation results obtained after removing the insignificant variables are given in Table 6.3.

Test	Jarque-Bera	White (no cross terms)	Ljung-Box (4 lags)
Test statistic (p-value)	0.1200 (0.9418)	51.8977 (0.0145)	3.1752 (0.5290)
Test	RESET (1 fitted term)	RESET (2 fitted terms)	RESET (3 fitted terms)
Test statistic (p-value)	0.0246 (0.8755)	9.5854 (0.0083)	12.0409 (0.0072)
Test	Serial LM (1 lag)	Serial LM (2 lags)	Serial LM (4 lags)
Test statistic (p-value)	0.0012 (0.9729)	0.8031 (0.6693)	4.5190 (0.3403)
Test	ARCH LM (1 lag)	ARCH LM (2 lags)	ARCH LM (4 lags)
Test statistic (p-value)	0.9555 (0.3283)	5.1675 (0.0755)	5.2638 (0.2613)

Table 6.4: Specification and diagnostic tests for linear Phillips curve (6.13)

Specification and diagnostic tests are performed to evaluate the estimated equation. The p-values of the Jarque-Bera test and the LM test of no error autocorrelation show that the null hypotheses of the normal distribution of the error term and of no error autocorrelation, respectively, cannot be rejected (Table 6.4). One can also see from Table 6.4 that there are no ARCH effects present in our model, while the CUSUM and CUSUMQ tests in Figures 6.5 and 6.6 detect only slight problems regarding parameter constancy.

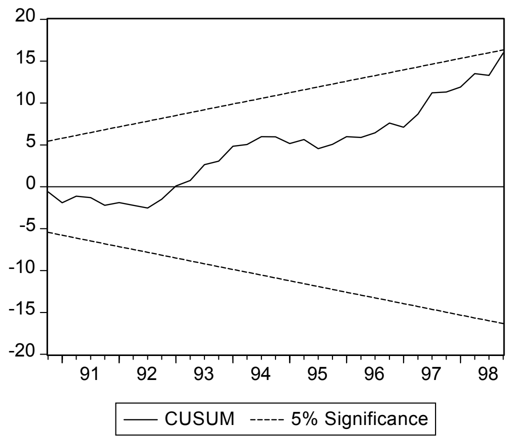

Figure 6.5: CUSUM test for linear Phillips curve (6.13)

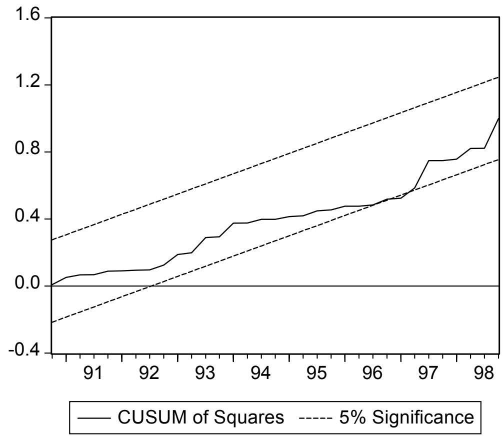

Figure 6.6: CUSUM of squares test for linear Phillips curve (6.13)

However, inspection of the properties of the estimated linear equation indicates misspecification problems (Table 6.4). The Ramsey RESET test with 2 and with 3 fitted terms rejects the null hypothesis of the normally distributed white noise errors. RESET is a general test for incorrect functional form based on an augmented auxiliary regression including powers of the estimated dependent variable (i.e. $\hat{y}^2, \hat{y}^3, \hat{y}^4, \ldots$) as explanatory variables. The White test statistic with the p-value of 0.01 rejects the null hypothesis of no heteroscedasticity. Actually, the White statistic tests three assumptions, due to the fact that under the null hypothesis the disturbances are supposed to be homoscedastic, independent of the regressors and that the linear specification of the model is supposed to be correct. Thus, if any of the three assumptions does not hold the test statistic gives significant values.

Recursive estimates of the linear Phillips curve are displayed in Figure 6.7. Several of the recursive estimates exhibit smooth change, thus signaling misspecification. Especially the coefficients $C(8)$, $C(5)$, and $C(6)$ of the forward looking price expectation variable (π_t^f), the $\Delta\pi_{t-8}$ and the $\Delta\pi_{t-12}$ variable, respectively, evolve gradually over time. Recall that the variable π_t^f was the one to exhibit the most coefficient variation also in case of the linear surrogate quantity equation (Figure 6.4).

Due to the misspecification problems indicated by the RESET test, White test and the recursive residuals, nonlinear specification will be considered for modeling the Phillips curve.

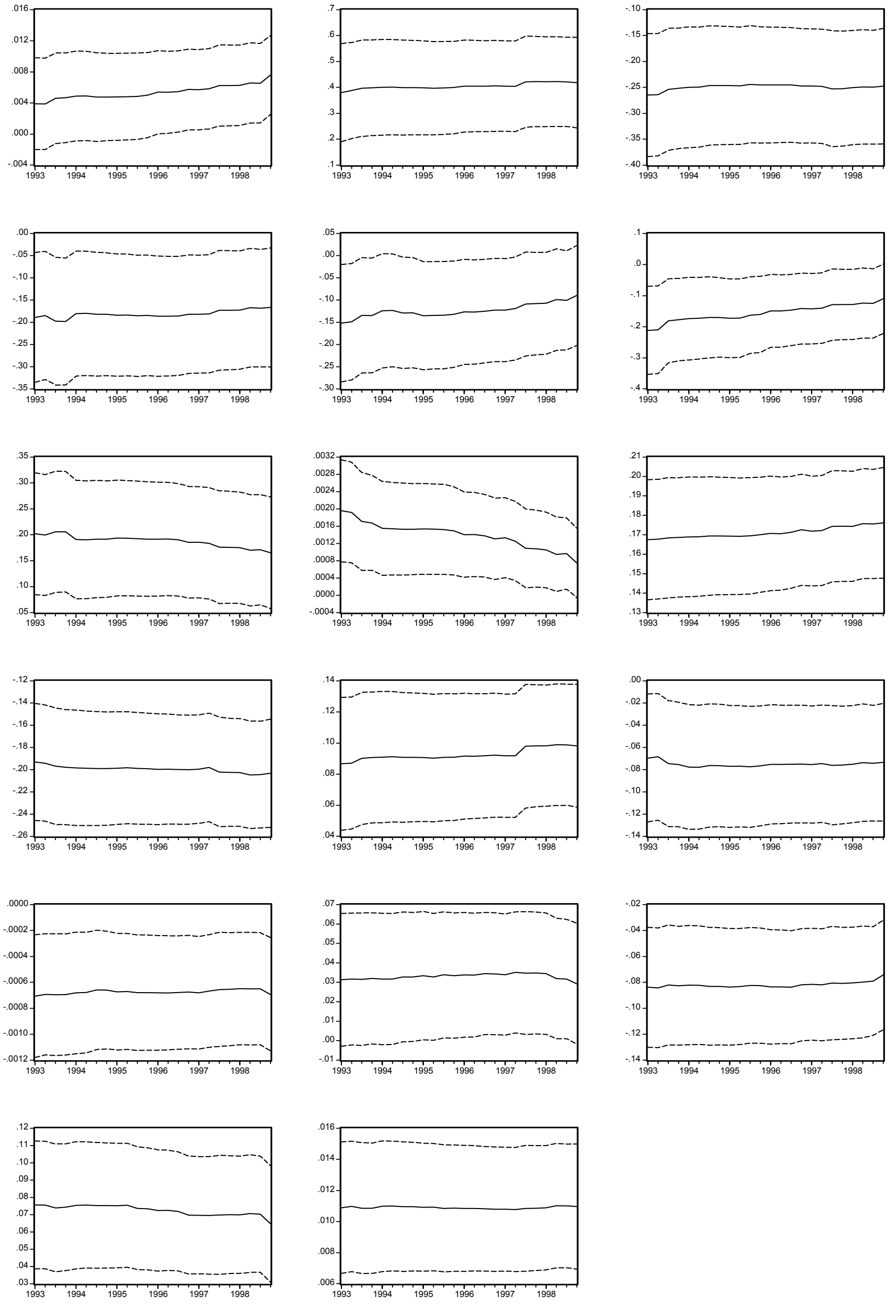

Figure 6.7: Recursive coefficients for linear Phillips curve (6.13) (following the estimated coefficients in Table 6.3 row-wise)

6.2.3 Okun's law

As already mentioned, Okun's law describes the short-run relationship between the the unemployment rate (u_t) and the output gap ($x_t - x_t^*$). x_t^* denotes the expected rate of real growth following the long-run trend. Proceeding as in Grant [27], four approaches for modeling x_t^* via the business cycle were applied: simple average, linear trend, the Hodrick-Prescott decomposition and the Beveridge-Nelson decomposition. Since simple average did not perform worse than the other three methods, it was chosen as the appropriate method.

Due to the unit root in the unemployment rate series, the first difference of the unemployment rate was employed as the dependent variable. Lagged variables u_t and Δu_t were included in the specification to account for the proper dynamics:

$$\Delta u_t = a + b(x_t - x^*) + c \cdot u_{t-1} + d(L)\Delta u_{t-1} + \varepsilon_t, \tag{6.14}$$

with the degree of the lag polynomial $d(L)$ restricted to 4. We found reasonable linear estimates of the Okun's law, which are given in Table 6.5.

OKUN'S LAW				
Dependent Variable: Δu_t				
Sample (adjusted): 1970:2 2000:3 (Observations: 122)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
$x_t - x^*$	-4.953883	0.805964	-6.146531	0.0000
const.	0.131311	0.031709	4.141159	0.0001
u_{t-1}	-0.017967	0.004797	-3.745722	0.0003
Δu_{t-1}	0.428198	0.068666	6.235993	0.0000
$dummy3$	0.636079	0.102657	6.196183	0.0000
$dummy4$	-0.433570	0.143116	-3.029498	0.0030
R^2 = 0.695, S.E. = 0.141, SSR = 2.298, AIC = -1.036				

Table 6.5: OLS results for linear Okun's law (6.14)

The estimated Okun's equation passed several of the specification and diagnostic tests, the results of which are given in Table 6.6. While there are no problems with the normal distribution of residuals, autocorrelation and heteroscedasticity, the Ramsey RESET test with 2 fitted terms indicates misspecification at the 5 % significance level. In addition, the ARCH

LM test with 2 and 4 lags detects autoregressive conditional heteroscedasticity, although this problem does not invalidate the properties of the OLS estimator. The CUSUM and CUSUMQ tests were used to investigate the parameter stability issue. According to the plots in Figures 6.8 and 6.9, the null hypothesis of parameter constancy cannot be rejected.

Test	Jarque-Bera	White (cross terms)	Ljung-Box (4 lags)
Test statistic (p-value)	0.9620 (0.6182)	9.0993 (0.6944)	3.1722 (0.5290)
Test	RESET (1 fitted term)	RESET (2 fitted terms)	RESET (3 fitted terms)
Test statistic (p-value)	0.0555 (0.8137)	6.0378 (0.0489)	6.6690 (0.0832)
Test	Serial LM (1 lag)	Serial LM (2 lags)	Serial LM (4 lags)
Test statistic (p-value)	0.3266 (0.5677)	0.4025 (0.8177)	0.4167 (0.9368)
Test	ARCH LM (1 lag)	ARCH LM (2 lags)	ARCH LM (4 lags)
Test statistic (p-value)	2.2328 (0.1351)	15.8188 (0.0004)	17.1305 (0.0018)

Table 6.6: Specification and diagnostic tests for linear Okun's law (6.14)

The recursive coefficients for the linear Okun's law are plotted in Figure 6.10. Four of the six recursive coefficient estimates exhibit significant variation over time, only the last two coefficients (of the two dummy variables) seem to be constant. Most pronounced are the changes in the coefficients $C(3)$ and $C(4)$ belonging to the lagged unemployment rate variable and the lagged first difference in the unemployment rate variable, respectively, around the year 1997. This indication of misspecification, together with the low p-value of the Ramsey RESET test, makes the Okun's law a suitable candidate for nonlinear specification.

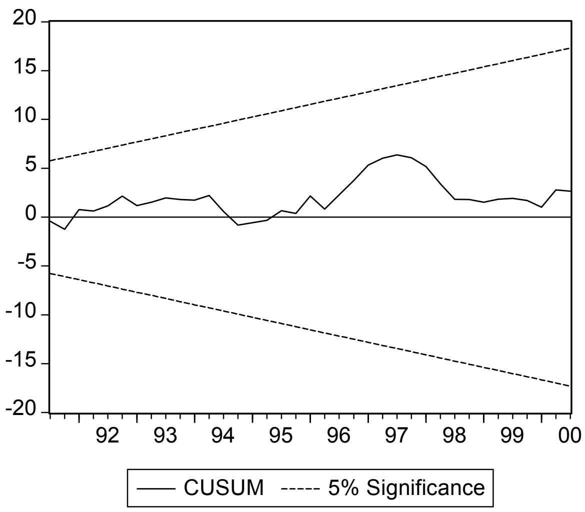

Figure 6.8: CUSUM test for linear Okun's law (6.14)

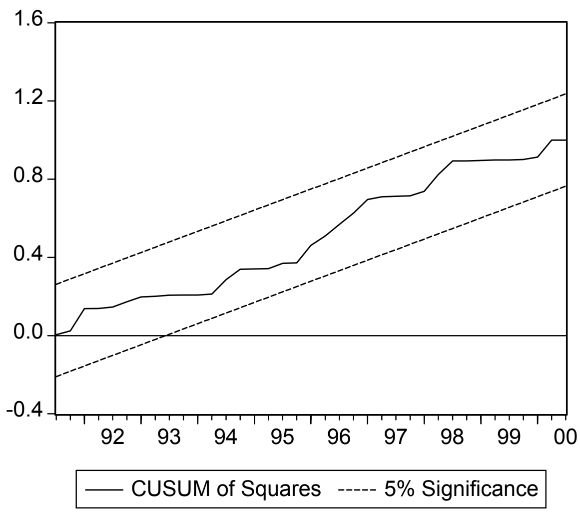

Figure 6.9: CUSUM of squares test for linear Okun's law (6.14)

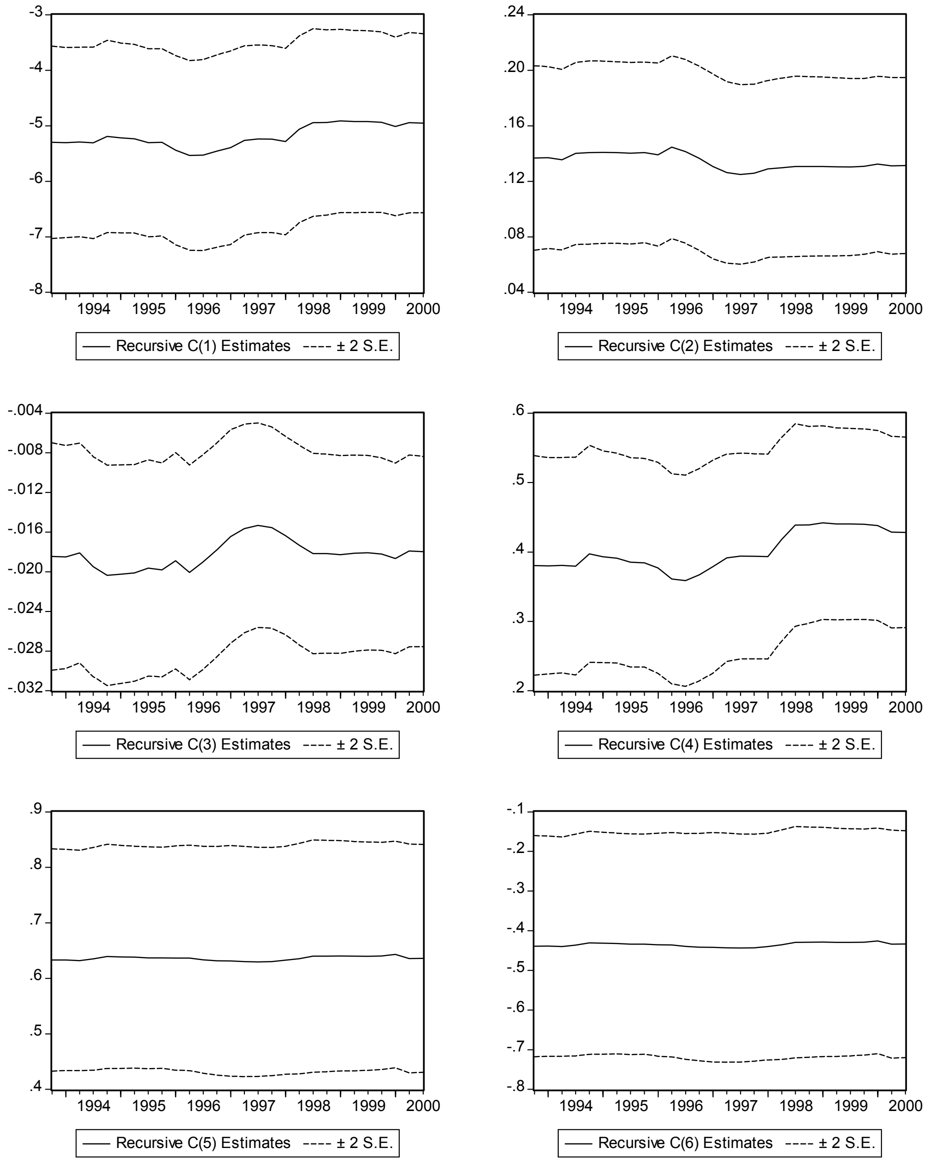

Figure 6.10: Recursive coefficients for linear Okun's law (6.14) (following the estimated coefficients in Table 6.5 row-wise)

It has to be emphasized that the dummy variables were introduced into each of the three equations to reduce the ARCH effects caused by the outliers in the years 1991 and 1992, when the German reunion took place. To sum up the results of this section, inspection of the properties of the estimated linear equations indicates in all cases problems with parameter stability or misspecification. CUSUM an CUSUMQ tests reveal structural breaks in the surrogate quantity equation in the period from 1991:3 till 1993:1, whereas the Ramsey RESET test with 2 fitted terms rejects the null hypothesis of NID errors for the Phillips curve and the Okun's law. Additionally, significant variation is revealed in some of the recursive coefficients of each equation. In order to improve specification we need to investigate the influence of nonlinearities, which we shall assume to be of the smooth transition kind.

Let us conclude with a remark regarding the estimation of our system. The linear model consisting of equations (6.11), (6.13) and (6.14) is a simultaneous system because of the endogenous variable u_t on the right-hand side of the Phillips curve given by equation (6.13). But this system is recursive, since it is possible to express sequentially each of the endogenous variables only in terms of predetermined variables. Due to problems with starting values, near singular moment matrices, no convergence of the nonlinear optimization algorithms and other numerical problems (when estimating the nonlinear model in section 6.3.1), each of the three equations in the system was estimated with single-equation methods also for the linear model. As the system is recursive, consistent estimates are obtained in this way.

6.3 Nonlinear model

Nonlinear money demand

Since representation of asymmetric reactions, structural changes and other phenomena of economic developments can be fruitfully investigated by nonlinear modeling techniques, the issue of a possible nonlinear money demand specification has been studied by several authors. Chen and Wu [11] show that employing the conventional linear cointegration approach in examining long-run money demand may not be appropriate after taking into account the existence of transaction costs. They provide evidence that deviations from equilibrium money demand follow an exponential smooth transition autoregressive process that is mean-reverting outside a given range and has a unit-root inside the range. Similarly, Sarno, Taylor and Peel [66] argue that several theoretical models of money demand imply nonlinear functional forms for the aggregate demand for money, characterized by smooth adjustment toward long-run equilibrium. Their paper proposes a nonlinear equilibrium correction model of U.S. money demand that is shown to be stable over the sample period from 1869 to 1997. The use of an exponential smooth transition regression model, with the lagged long-run equilibrium error acting as the transition variable, implies faster adjustment toward equilibrium, the greater the absolute size of the deviation from equilibrium. In a similar study, Sarno [65] presents a stable empirical model for the demand for narrow money in Italy using annual data spanning from Italian unification in 1861 through to 1991. A nonlinear functional form of the aggregate demand for money is characterized by smooth adjustment towards long-run equilibrium, again achieved by estimating a nonlinear error correction model in the form of an exponential smooth transition regression.

Several authors studied the money demand equation during hyperinflation. According to the findings of Petrović and Mladenović [61], the modified money demand in Yugoslavia in the period from 1992 till 1994 is nonlinear with decreasing semielasticity of money demand. Tallman, Tang and Wang [70], on the other hand, employed the data from the Chinese hyperinflationary episode, expanding the standard Caganian money demand function to include both anticipated inflation and relative price effects in a nonlinear fashion. Their empirical findings suggest that conventional econometric investigations of money demand during hyperinflation overlook important nonlinear interactions between real and monetary activities and, hence, underestimate the welfare costs of hyperinflation.

Nonlinear Phillips curve

Substantial theoretical and empirical evidence can be found in the literature suggesting nonlinearity in the output-inflation relationship, namely a nonlinear Phillips curve. Dolado, Ramon and Naveira [17] investigate the implications of a nonlinear Phillips curve for the derivation of optimal monetary policy rules. They show that combined with a quadratic loss function, the optimal policy is also nonlinear, with the policy-maker increasing interest rates by a larger amount when inflation or output are above target than the amount it will reduce them when they are below target. The main prediction of their model is that such a source of nonlinearity leads to the inclusion of the interaction between expected inflation and the output gap in an otherwise linear Taylor rule. The model of Schaling [67] features a convex Phillips curve, in that positive deviations of aggregate demand from potential are more inflationary than negative deviations are disinflationary. Corrado and Holly [14] consider the performance of optimal policy rules when the underlying relationship between inflation and the output gap may be nonlinear. In particular, if the inflation-output trade-off exhibits nonlinearities this will impart a bias to inflation when a linear rule is used. To correct this bias they propose a piecewise linear rule, which can be thought of as an approximation to the nonlinear rule of Schaling [67].

Hooker [36] identifies a structural break in core US inflation Phillips curves such that oil prices contributed substantially before 1981, but since that time pass-through has been negligible. In the framework of a Keynesian monetary macro model, Chiarella et al. [12] study implications of kinked Phillips curves and alternative monetary policy rules. Nobay and Peel [56] analyze optimal discretionary monetary policy under a non-linear Phillips curve. It is shown that the results are in marked contrast to conventional results that are drawn from the linear paradigm. Specifically, there exists a deflation bias in expected output, while the inflation bias cannot be signed. Collard and Juillard [13] propose to apply to the simulation of general nonlinear rational-expectation models a method where the expectation functions are approximated through a higher-order Taylor expansion. Their macroeconomic model features a nonlinear Phillips curve. Tambakis [71] argues that recent theoretical and empirical work has cast doubt on the hypotheses of a linear Phillips curve and a symmetric quadratic loss function underlying traditional thinking on monetary policy. In his paper he studies the one-period optimal monetary policy problem under an asymmetric loss function corresponding to the "opportunistic approach" to disinflation and a convex Phillips curve.

Mayes and Viren [53] highlight the implications for a single monetary policy when key economic relationships are nonlinear or asymmetric at a disaggregate level. Using data for the EU and OECD countries they show that there are considerable non-linearities and asymmetries in the Phillips and Okun curves. High unemployment has relatively limited effect in pulling inflation down while low unemployment can be much more effective in driving it up.

To accommodate potentially important departure from linearity of the Phillips curve, Huh [37] employs a vector autoregression (VAR) model of output, inflation, and the terms of trade augmented with logistic smooth transition autoregression specifications. Empirical results indicate that the model captures the nonlinear features present in the data well. Based on this nonlinear approximation, the output costs for reducing inflation are found to vary, depending critically on the state of the economy, the size of intended inflation change, and whether policymakers seek to disinflate or prevent inflation from rising. This implies that inferences based on the conventional linear Phillips curve may provide misleading signals about the cost of lowering inflation and thus the appropriate policy stance. Böhm [6] also employs the smooth transition regression modeling approach. In a formulation of an inflation equation for Austria, which includes the demand and supply features, he explores the capacity of STR models to improve upon specification. The nonlinearities and asymmetries are found to be relevant ingredients in the Austrian inflation equation and the change in the unemployment rate is shown to have a larger impact on inflation during periods of high volatility of price increases.

Nonlinear Okun‘s law

While the linear relationship between output and unemployment rate in the United States was established empirically by Okun, Prachowny [63] provided theoretical derivation of the relation in a special case. Under the assumptions that the aggregate production function is of a Cobb - Douglas type and that the capital stock and a disembodied technology factor are always at their long - run levels, Prachowny established a log linear relationship between the output gap and capacity utilization gap, labour supply gap and hours worked gap. Weber and West [79] used the Box - Cox transformation, which allows empirical testing of the log linear Okun's law specification against more general alternatives. They found strong support in favour of the Prachowny's log linear functional form of the Okun's law. The issue of stability of the Okun's law relationship has been discussed extensively by several authors. Blanchard [5] claims that the stability of the Okun's coefficient has decreased with time. The effect of the output change on unemployment is supposed to be stronger due to the intense international competition, less legal protection for the workers and reduced labour hoarding. Sögner and Stiassny [68] use Baysian methods to test for discrete structural breaks in the Okun's law and Kalman filter to check for continuous parameter changes. 15 OECD countries are included in their study. The first approach does not detect any structural breaks, whereas the results of the second approach imply continuous parameter changes for 10 of the countries. The relationships between output and labour demand and labour supply, respectively, are also discussed in the paper. The authors conclude that for most countries the change in the Okun's coefficient results mainly from an increased reaction of employment to GDP change. A nonlinear relationship between cyclical unemployment and cyclical output is proposed by Cuaresma [15]. For the US data, the linear specification is strongly rejected in favor of a piecewise linear specification. The estimated Okun's coefficient is significantly higher for expansions than for recessions, implying that output changes cause asymmetric and regime dependent changes in the unemployment rate. Additionally, unemployment shocks tend to be more persistent in times of expansion. The findings of Mayes and Viren [53] for EU and OECD countries are similar. Asymmetry is built into the Okun's law with the help of the threshold model and the error correction mechanism, which enables regime dependent correction paths. Most of the countries included in the study exhibit asymmetric relationship between unemployment rate and change in output.

6.3.1 Testing linearity against STR

Each of the three equations is now subjected to linearity tests. In every case, all significant explanatory variables from the estimated linear equation are considered as suitable candidates for a transition variable, with the addition of the time trend. The OLS estimates of the linear surrogate quantity equation, linear Phillips curve and linear Okun's law can be found in Tables 6.1, 6.3 and 6.5, respectively.

Surrogate quantity equation

The results of the linearity tests for the surrogate quantity equation are given in the second column of Table 6.7. The variable x_{t-4} does not exhibit potential for nonlinear specification,whereas the test is strongly rejected for all other six candidates. As explained in subsection 3.2.1, the variable m_{t-4} with the strongest rejection of the null hypothesis of linearity, i.e. with the lowest p-value, is selected for the transition variable.

The next step in the modeling process consists of choosing the type of the transition function. The decision rule is based on a sequence of nested hypotheses, which are tested with F-type tests named F4, F3, and F2 (see subsection 3.2.2 for details). Basically, the aim of these tests is to determine the degree of the polynomial in auxiliary regression (3.22) obtained after the transition function in the STR model is replaced by its Taylor approximation around $\gamma = 0$. Following Teräsvirta [73], the LSTR1 transition function is chosen, as the F2 test has the lowest p-value in case of the previously selected transition variable m_{t-4} (Table 6.7).

SURROGATE QUANTITY EQUATION				
Variable	F-test	F4-test	F3-test	F2-test
$ttrend$	2.3383 (0.0028)	2.6015 (0.0174)	2.6158 (0.0164)	1.4620 (0.1894)
x_t	2.1617 (0.0062)	1.5166 (0.1722)	1.7743 (0.1015)	3.1389 (0.0048)
x_{t-4}	1.4783 (0.1034)	0.7533 (0.6277)	1.3475 (0.2370)	2.5564 (0.0182)
$\Delta\pi_{t-1}$	2.3838 (0.0023)	1.7722 (0.1031)	2.9608 (0.0075)	2.1552 (0.0445)
m_{t-1}	2.5023 (0.0014)	1.1519 (0.3389)	3.0606 (0.0060)	3.1299 (0.0049)
m_{t-4}	2.5366 (0.0012)	3.0211 (0.0069)	0.8138 (0.5780)	3.5351 (0.0019)
π_t^f	1.8201 (0.0266)	2.4554 (0.0240)	1.2178 (0.3008)	1.7021 (0.1169)
df	22, 87	7, 87	7, 94	7, 101

Table 6.7: F-values (and p-values) of the linearity tests for the surrogate quantity equation

Phillips curve

From the second column of Table 6.8 with the linearity test results for the Phillips curve we obtain a set of nine suggested transition variables (with p-values below 0.05), from which we choose the variable with the strongest rejection of linearity. This is the variable m_{t-2} in our case. Note that lagged money growth was selected to be the transition variable also for the surrogate quantity equation, only that it is now lagged 2 times.

PHILLIPS CURVE				
Variable	F-test	F4-test	F3-test	F2-test
$ttrend$	1.8274	2.1217	1.1073	1.6001
	(0.0174)	(0.0217)	(0.3667)	(0.0869)
$\Delta\pi_{t-1}$	2.0394	1.4443	2.4285	1.4994
	(0.0064)	(0.1592)	(0.0060)	(0.1196)
π_{t-1}	1.6366	1.5258	0.9073	2.2298
	(0.0419)	(0.1270)	(0.5643)	(0.0098)
π_{t-4}	1.2605	0.7590	1.4331	1.7142
	(0.2073)	(0.7214)	(0.1535)	(0.0597)
$\Delta\pi_{t-8}$	0.8453	0.4812	1.1130	1.2651
	(0.7219)	(0.9453)	(0.3617)	(0.2393)
$\Delta\pi_{t-12}$	1.3384	1.2311	1.3955	1.2288
	(0.1524)	(0.2781)	(0.1712)	(0.2644)
π_{t-5}	1.3137	0.9873	1.9154	0.9602
	(0.1683)	(0.4842)	(0.0340)	(0.5065)
π_t^f	1.8682	1.3288	1.7354	2.0472
	(0.0144)	(0.2168)	(0.0609)	(0.0189)
π_t^0	1.6030	1.2918	1.2029	2.0969
	(0.0488)	(0.2386)	(0.2893)	(0.0158)
π_{t-1}^0	1.3350	1.2738	0.6912	2.0742
	(0.1545)	(0.2498)	(0.7926)	(0.0172)
π_{t-2}^0	1.3113	1.1995	0.7013	2.1086
	(0.1699)	(0.3006)	(0.7827)	(0.0152)
λ_{t-2}	1.0655	0.9075	1.2097	1.1526
	(0.4110)	(0.5654)	(0.2842)	(0.3233)
u_t	1.8317	1.8731	1.1664	1.8654
	(0.0170)	(0.0461)	(0.3173)	(0.0357)
m_{t-2}	2.4265	3.0185	0.9951	1.9842
	(0.0010)	(0.0014)	(0.4727)	(0.0236)
m_{t-3}	1.9963	1.6067	1.9518	1.6771
	(0.0079)	(0.1009)	(0.0301)	(0.0675)
m_{t-4}	1.9241	1.6857	2.2049	1.1442
	(0.0110)	(0.0803)	(0.0129)	(0.3302)
df	49, 51	16, 51	16, 67	16, 83

Table 6.8: F-values (and p-values) of the linearity tests for the Phillips curve

A comparison of the F4, F3 and F2 test statistics for the transition variable m_{t-2} in Table 6.8 shows that F4 yields the lowest p-value, thus indicating the LSTR1 transition function.

Okun‘s law

When it comes to the Okun‘s law, the results are unexpected. The null hypothesis of linearity is not rejected for any candidate for the transition variable (second column of Table 6.9), not even at the 10 % significance level.

OKUN‘S LAW				
Variable	F-test	F4-test	F3-test	F2-test
ttrend	1.3097	2.4543	0.9118	0.8000
	(0.2193)	(0.0504)	(0.4599)	(0.5277)
$x_t - x^*$	1.5387	2.0633	1.9709	0.8253
	(0.1161)	(0.0910)	(0.1042)	(0.5118)
u_{t-1}	0.8491	0.2426	1.4665	1.1052
	(0.6079)	(0.9135)	(0.2175)	(0.3578)
Δu_{t-1}	1.2270	0.6275	1.8459	1.5155
	(0.2712)	(0.6440)	(0.1254)	(0.2025)
df	13, 103	4, 103	4, 107	4, 111

Table 6.9: F-values (and p-values) of the linearity tests for the Okun‘s law

Because of the low p-value of the Ramsey RESET test and the variation in the recursive coefficients of the linear Okun‘s law, the linearity test was expected to be rejected as well. We also tried to estimate Okun‘s law using seasonally unadjusted data for West Germany. In order to deal with seasonality we have chosen to work with the fourth difference of the unemployment rate u_t as dependent variable. After several attempts of specifying a model linear in the output gap, the squared gap variable was also employed. The outcome was just the opposite as for seasonally adjusted data. Namely, linearity was rejected for 4 of the 5 possible transition variables. The results of the estimation are given in section 6.5. The comparison of both attempts at specifying and estimating Okun‘s law points out the sensitivity of the nonlinear modeling process and shows that the final outcome can be influenced by several factors.

6.3.2 Estimates of the nonlinear equations

When estimating the STR specifications of the surrogate quantity equation and the Phillips curve, all of the significant variables from the preliminary linear model with the estimates given in Tables 6.1 and 6.3 were considered as explanatory variables. Due to occasional problems with convergence of the nonlinear optimization procedure some experimentation to find appropriate starting values was required. The final set of estimates obtained for the parameters of the nonlinear surrogate quantity equation and the nonlinear Phillips curve can be found in equations (6.15) and (6.16).

Nonlinear surrogate quantity equation

Removing the insignificant variables from the nonlinear surrogate quantity equation yields

$$\begin{aligned} m_t &= \underset{(0.0039)}{0.0112} + \underset{(0.1194)}{0.2213} \cdot x_t + \underset{(0.0878)}{0.1766} \cdot x_{t-4} - \underset{(0.2259)}{0.6599} \cdot \Delta\pi_{t-1} + \\ &+ \underset{(0.0500)}{0.9274} \cdot m_{t-1} + \underset{(0.0166)}{0.0920} \cdot dummy1 + \left(\underset{(0.0793)}{-0.2096} \cdot m_{t-1} - \underset{(0.0022)}{0.0104} \cdot \pi_t^f \right) \cdot \\ &\cdot \left[1 + exp\{\underset{(1.6579)}{-4.1174}(m_{t-4} - \underset{(0.0101)}{0.0572})/0.0571\}\right]^{-1}. \end{aligned} \tag{6.15}$$

The signs and the magnitudes of the coefficients are in accordance with the underlying economic theory. The γ value of approximately 4.12 indicates moderate speed of transition between the two extreme regimes, as shown in Figure 6.11, where the transition function is plotted against the transition variable.

The strongly nonlinear behavior implied by our empirical estimates of the nonlinear surrogate quantity equation is made clear by the plot of the estimated transition function against the transition variable given in Figure 6.12. The plot displays significant variation in the values of the transition function and shows that also the extreme regime with $G = 1$ is achieved.

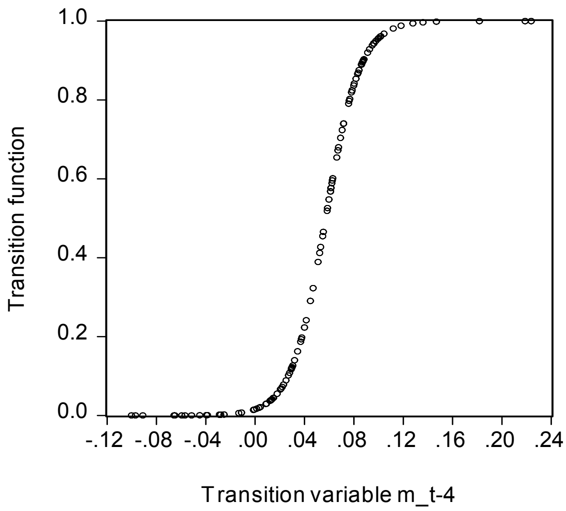

Figure 6.11: Scatter plot of the transition function against the transition variable for surrogate quantity equation (6.15)

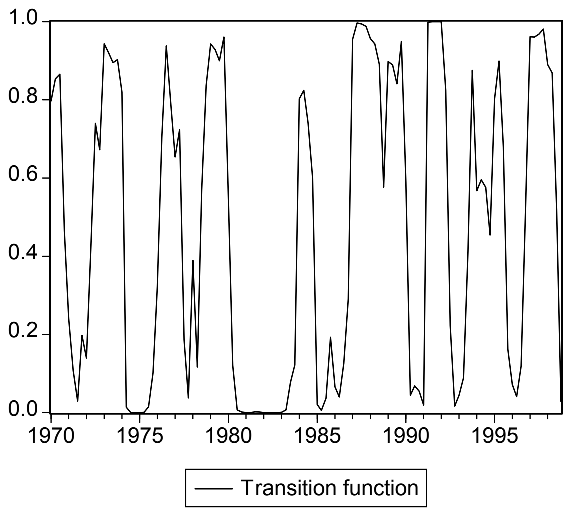

Figure 6.12: Transition function for surrogate quantity equation (6.15)

Next, the estimated equation is submitted to the specification and diagnostic tests. As shown by results given in Tables 6.10, 6.11 and 6.12, the equation displays white noise residuals and shows no evidence of ARCH effects and error autocorrelation.

Jarque-Bera test p-value	0.0799
Ljung-Box p-value (8 lags)	0.5679
AIC	-7.6044
R^2	0.8699
S. E. of residuals	0.0214
Ratio var(nonlin)/var(lin)	0.7231
No. of observations	116

Table 6.10: Specification and diagnostic tests for nonlinear surrogate quantity equation (6.15)

Lag	Autocorrelation	McLeod-Li p-value	ARCH LM p-value
1	0.1109	0.1891	0.1984
2	-0.0171	0.1031	0.1417
3	-0.0094	0.1876	0.2749
4	-0.1527	0.3135	0.4276
5	0.0194	0.3359	0.4489
6	0.0521	0.0637	0.1040
7	0.1066	0.0632	0.1555
8	-0.0891	0.0560	0.2247

Table 6.11: McLeod-Li and ARCH tests for nonlinear surrogate quantity equation (6.15)

Another key assumption of STR estimation is the parameter constancy. The LM1, LM2 and LM3 tests described in section 3.4.3 were applied to investigate the stability of parameters. As one can see from the results in Tables 6.13 and 6.14, the LM3 test yields a p-value just below 0.05 for the lagged money growth variable m_{t-1} in the nonlinear part of the surrogate quantity equation and thus points to a problem with the stability of this coefficient estimate. Recall that the LM3 test was developed for detecting time-dependent parameters following a non-monotonous function under the

No. of lags	F-statistics	Deg. of freedom	p-value
1	1.3505	1, 105	0.2478
2	0.6704	2, 104	0.5137
3	0.4505	3, 103	0.7175
4	1.0722	4, 102	0.3743
5	0.8671	5, 101	0.5061
6	0.7279	6, 100	0.6282
7	0.7774	7, 99	0.6077
8	0.9154	8, 98	0.5073

Table 6.12: LM test of no remaining error autocorrelation for nonlinear surrogate quantity equation (6.15)

alternative hypothesis. Except for m_{t-1} in the nonlinear part of the equation, the null hypothesis of parameter constancy cannot be rejected.

Variable: *const*				Variable: x_t			
Test	F-statistic	df	p-value	Test	F-statistic	df	p-value
LM3	2.5699	3, 103	0.0583	LM3	1.4952	3, 103	0.2203
LM2	0.3091	2, 104	0.7348	LM2	0.4695	2, 104	0.6266
LM1	0.1293	1, 105	0.7199	LM1	0.1168	1, 105	0.7332
Variable: x_{t-4}				Variable: $\Delta\pi_{t-1}$			
Test	F-statistic	df	p-value	Test	F-statistic	df	p-value
LM3	1.0553	3, 103	0.3715	LM3	1.2018	3, 103	0.3129
LM2	0.6741	2, 104	0.5118	LM2	0.4556	2, 104	0.6353
LM1	0.6241	1, 105	0.4313	LM1	0.4928	1, 105	0.4842
Variable: m_{t-1}							
Test	F-statistic	df	p-value				
LM3	1.0791	3, 103	0.3614				
LM2	0.1179	2, 104	0.8889				
LM1	0.0533	1, 105	0.8177				

Table 6.13: Parameter constancy test for the parameters in the linear part of surrogate quantity equation (6.15)

To sum up, the null hypothesis of linearity tested against the alternative

Variable: m_{t-1}				Variable: π_t^f			
Test	F-statistic	df	p-value	Test	F-statistic	df	p-value
LM3	2.7304	3, 103	0.0477	LM3	1.5439	3, 103	0.2077
LM2	1.2224	2, 104	0.2987	LM2	2.3334	2, 104	0.1020
LM1	1.6606	3, 105	0.2004	LM1	0.0822	3, 105	0.7750

Table 6.14: Parameter constancy test for the parameters in the nonlinear part of surrogate quantity equation (6.15)

smooth transition regression specification had to be rejected for every possible transition variable with the exception of x_{t-4}, which confirms our intuition that the linear relationship of the surrogate quantity equation (estimated as money demand equation) can be improved by consideration of regime changes. Our results suggest that the failure to allow for nonlinear dynamics may help explain the difficulty of much empirical research in obtaining stable money demand equations.

Nonlinear Phillips curve

The estimated nonlinear Phillips curve is given by the next equation:

$$\begin{aligned}
\Delta\pi_t \; = \; & \underset{(0.0122)}{0.1874} \cdot \pi_t^0 + \underset{(0.0240)}{0.0548} \cdot \pi_{t-5} - \underset{(0.0233)}{0.0545} \cdot \lambda_{t-2} - \\
- \; & \underset{(0.0154)}{0.0582} \cdot m_{t-3} + \underset{(0.0133)}{0.0894} \cdot \pi_{t-2}^0 + \underset{(0.0150)}{0.0607} \cdot m_{t-4} + \underset{(0.0023)}{0.0118} + \\
+ \; & \underset{(0.0018)}{0.0129} \cdot dummy2 + \underset{(0.0291)}{0.4132} \cdot \Delta\pi_{t-1} - \underset{(0.0478)}{0.3286} \cdot \pi_{t-1} - \\
- \; & \underset{(0.0591)}{0.1478} \cdot \Delta\pi_{t-12} - \underset{(0.0188)}{0.1867} \cdot \pi_{t-1}^0 - \underset{(0.0002)}{0.0012} u_t + \\
+ \; & (\underset{(0.0863)}{0.1768} \cdot \Delta\pi_{t-12} - \underset{(0.0111)}{0.0228} \cdot \pi_{t-1}^0 + \underset{(0.0002)}{0.0004} \cdot u_t) \cdot \\
\cdot \; & [1 + exp\{-42.8009(m_{t-2} - \underset{(0.0057)}{0.0599})/\underset{(0.0027)}{0.0569}\}]^{-1}.
\end{aligned} \tag{6.16}$$

The value of the slope parameter γ is approximately 42.8, which is much higher than in case of the surrogate quantity equation. One can see from

Figure 6.13 that the transition function changes abruptly from 0 to 1 when the transition variable m_{t-2} approaches the threshold value of 0.06. Thus, the estimated smooth transition regression model is close to switching regression.

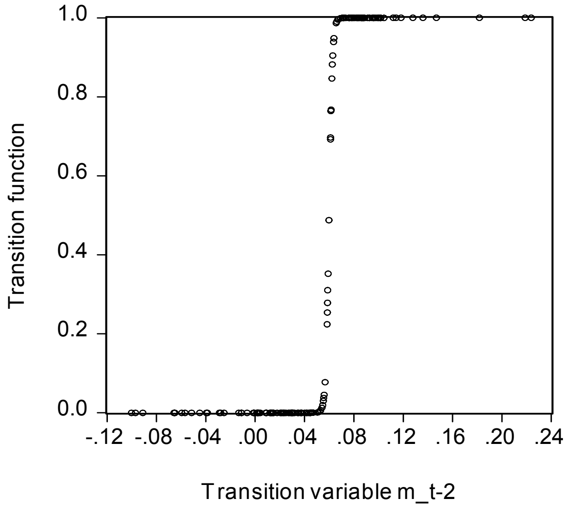

Figure 6.13: Scatter plot of the transition function against the transition variable for Phillips curve (6.16)

By plotting the transition function against time, the frequent changes between the two extreme regimes become apparent (Figure 6.14). Both the values of 0 and the values near 1 are attained several times during the sample period.

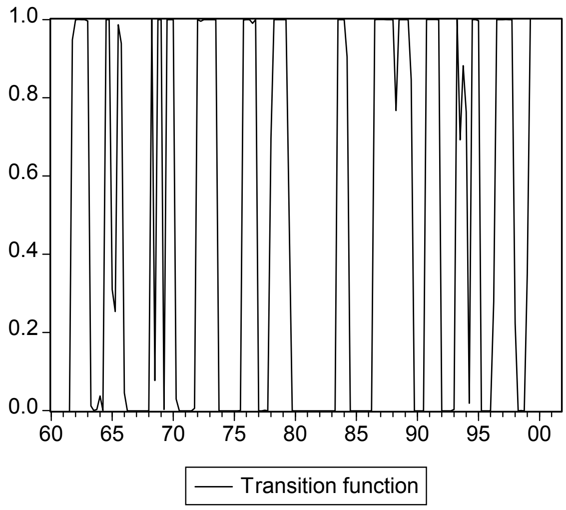

Figure 6.14: Transition function for Phillips curve (6.16)

The estimated nonlinear Phillips curve passed the diagnostic checking satisfactorily. The p-values of the Jarque-Bera test and the test of no remaining error autocorrelation show that the null hypotheses of the normal distribution of the error term and of no error autocorrelation, respectively, cannot be rejected. The tests of parameter constancy detects no problems and there are also no ARCH effects present in our models. The results are given in Tables 6.15, 6.16, 6.17 and 6.18.

Jarque-Bera test p-value	0.5817
Ljung-Box p-value (8 lags)	0.2027
AIC	-10.8753
R^2	0.8167
S. E. of residuals	0.0041
Ratio var(nonlin)/var(lin)	0.9441
No. of observations	116

Table 6.15: Specification and diagnostic tests for nonlinear Phillips curve (6.16)

Lag	Autocorrelation	McLeod-Li p-value	ARCH LM p-value
1	0.0019	0.5365	0.5377
2	-0.0056	0.1451	0.1570
3	0.11529	0.2487	0.2830
4	-0.1986	0.3926	0.4393
5	0.0156	0.3507	0.3451
6	0.1529	0.4858	0.4854
7	-0.0017	0.5706	0.6130
8	-0.1367	0.6437	0.7090

Table 6.16: McLeod-Li and ARCH tests for nonlinear Phillips curve (6.16)

No. of lags	F-statistics	Deg. of freedom	p-value
1	0.00004	1, 97	0.9949
2	0.00878	2, 96	0.9913
3	0.79609	3, 95	0.4990
4	1.58683	4, 94	0.1842
5	1.25602	5, 93	0.2897
6	1.42658	6, 92	0.2129
7	1.27978	7, 91	0.2692
8	1.56149	8, 90	0.1477

Table 6.17: LM test of no remaining error autocorrelation for nonlinear Phillips curve (6.16)

The nonlinear Phillips curve proposed here yields a slight improvement in R^2 relative to the best-fitting linear equation in Table 6.3 and appears to be superior in several respects, also passing the battery of diagnostic tests and displaying parameter constancy despite the number of fundamental changes characterizing the West German history over our sample period.

6.4 Simulation and interpretation

We have thus obtained the final estimates of our nonlinear model. Since the null of linearity was not rejected for any of the possible transition variables

Variables in the linear part				Variables in the nonlinear part			
Test	F-statistic	df	p-value	Test	F-statistic	df	p-value
LM3	2.1430	36, 62	0.0712	LM3	0.5490	9, 89	0.8347
LM2	1.5388	24, 74	0.0821	LM2	0.7435	6, 92	0.6161
LM1	1.6825	12, 86	0.0848	LM1	0.3648	3, 95	0.7785

Table 6.18: Parameter constancy test for the parameters in the linear and nonlinear part of Phillips curve (6.16)

of the Okun's law, linear Okun's law with estimates given in Table 6.5 is added to nonlinear surrogate quantity equation (6.15) and nonlinear Phillips curve (6.16) to complete our system. The comparison of the linear system estimates given in Tables 6.1, 6.3 and 6.5 with the estimated nonlinear model reveals a slight increase in the explanatory power and a slight decrease in the standard error of regression for the surrogate quantity equation and for the Phillips curve.

The estimated nonlinear system is also inspected for its dynamic properties. Contrary to estimation, the money stock is considered a policy instrument to influence the inflationary process as well as the unemployment rate. Additionally, the reaction of a change in oil prices, in the forward looking expected inflation rate and to a productivity shock is also evaluated. In the latter cases, reactions turn out to be rather symmetric, while shocks in the money stock exhibit significant asymmetries. The graphs in Figure 6.15 and Figure 6.16 show the impact of a unit shock of one standard deviation of the money stock to inflation and unemployment.

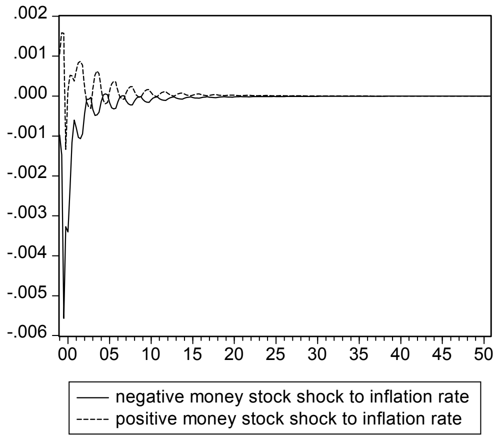

Figure 6.15: Effect of an initial unit shock of the money stock to inflation

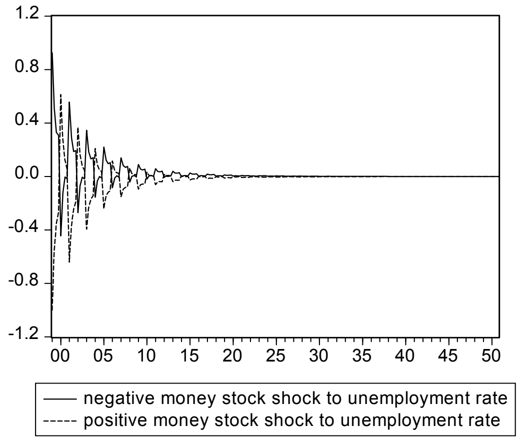

Figure 6.16: Effect of an initial unit shock of the money stock to unemployment

Maintained unit shock of one standard deviation of the money stock creates a permanent increase in inflation by 2.4 % points and a decrease in unemployment rate by 8.2 % points. Also in this case the asymmetric reaction to monetary policy is obvious (Figures 6.17 and 6.18). This again shows that working with a nonlinear model generates a result that is known in practice, but not achieved by linear models.

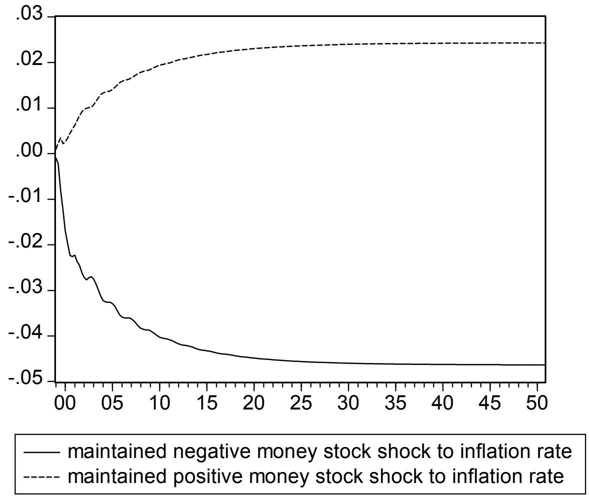

Figure 6.17: Effect of a maintained unit shock of the money stock to inflation

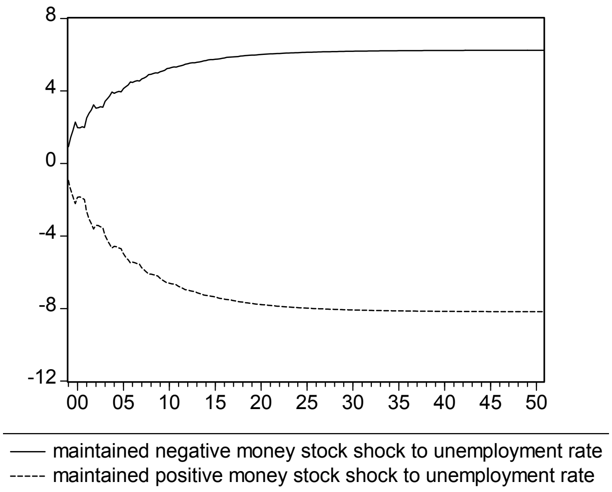

Figure 6.18: Effect of a maintained unit shock of the money stock to unemployment

6.5 Okun‘s law with seasonally unadjusted data

In this case, we used seasonally unadjusted quarterly data for West Germany for the period from 1970 till 1998. After several attempts of specifying a model linear in the output gap, the following estimation result including the squared gap was obtained. In order to deal with seasonality we have chosen to work with

$$v_t = u_t - u_{t-4} \tag{6.17}$$

(fourth difference of the unemployment rate u_t) as dependent variable. Additionally, it has proved useful to apply a difference transformation

$$\Delta v = v_t - v_{t-1}$$

when searching for the appropriate dynamics. In the table below, gap stands for the difference $x_t - x^*$, with the expected rate of real growth x^* set constant and equal to the arithmetic mean of x. The estimation results are given in Table 6.19.

Dependent variable: Δv				
Variable	Coefficient	Std. Error	t-value	p-value
gap_t	-0.046370	0.009330	-4.970098	0.0000
gap_t^2	0.008583	0.002319	3.700537	0.0003
$const$	-0.019394	0.022577	-0.858988	0.3921
v_{t-1}	-0.196022	0.031227	-6.277342	0.0000
Δv_{t-1}	0.424695	0.060624	7.005416	0.0000
Δv_{t-3}	0.178839	0.074396	2.403888	0.0178
Δv_{t-4}	-0.188553	0.070806	-2.662937	0.0088
$dummy1_t$	-0.666660	0.199169	-3.347206	0.0011
$dummy2_t$	0.541498	0.119728	4.522714	0.0000
T=127, R^2 = 0.6276, S.E.= 0.1946, AIC= -0.3675				

Table 6.19: Estimation results for Okun’s law

The obtained linear model proved satisfactory after being tested for normality, autocorrelations, ARCH effects and constancy of coefficients, but the Ramsey RESET test and the White test indicate problems due to misspecification. Test results can be found in Table 6.20 and in Figure 6.19.

Test	Jarque-Bera	Breusch-Godfrey (6 lags)	Ljung-Box (6 lags)
Test statistic (p-value)	3.3828 (0.1843)	1.6028 (0.1529)	4.8098 (0.5680)
Test	ARCH LM (4 lags)	Ramsey RESET (2 fitted terms)	White (cross terms)
Test statistic (p-value)	3.5073 (0.4768)	7.0194 (0.0299)	61.0637 (0.0007)

Table 6.20: Specification and diagnostic tests

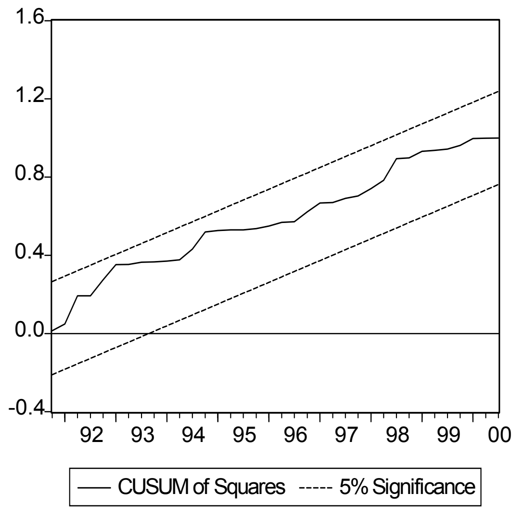

Figure 6.19: Cusum of squares

Next, the linearity test is performed for each of the possible transition variables. The variable Δv_{t-1} is chosen for the transition variable, as it has the lowest p-value (second column of Table 6.21). The p-values of the tests F4, F3 and F2 (also given Table 6.21) are calculated and compared in order to decide upon the type of the transition function. With the smallest one being the p-value of F3, we select the ESTR model. Alternatively, LSTR2 model could be chosen and the hypothesis $c_1 = c_2$ tested to see which of the models ESTR and LSTR2 is more appropriate.

Variable	F-test	F4-test	F3-test	F2-test
$ttrend$	1.3656 (0.1553)	1.8467 (0.0869)	0.8566 (0.5435)	1.2839 (0.2648)
gap^2	2.1836 (0.0055)	0.8733 (0.5306)	2.3991 (0.0257)	3.0576 (0.0056)
v_{t-1}	1.8242 (0.0261)	0.8555 (0.5446)	2.4293 (0.0240)	2.0486 (0.0551)
Δv_{t-1}	4.0094 (0.0000)	3.0765 (0.0057)	4.3349 (0.0003)	2.9076 (0.0079)
Δv_{t-3}	2.2833 (0.0036)	2.5188 (0.0201)	2.5255 (0.0194)	1.2808 (0.2663)
Δv_{t-4}	1.7681 (0.0330)	2.7082 (0.0131)	1.8583 (0.0840)	0.4459 (0.8711)
df	21, 97	7, 97	7, 104	7, 111

Table 6.21: F-values (and p-values) of the linearity test

After eliminating the insignificant variables, one obtains the estimated coefficients as shown in equation (6.18):

$$\begin{aligned}
\Delta v_t \;=\;& \underset{(0.0341)}{-0.1099} \cdot v_{t-1} + \underset{(0.0599)}{0.4727} \cdot \Delta v_{t-1} + \underset{(0.0657)}{0.1212} \cdot \Delta v_{t-3} - \\
-\;& \underset{(0.0084)}{0.0422} \cdot gap_t - \underset{(0.1683)}{0.7308} \cdot dummy1_t + \underset{(0.1039)}{0.6099} \cdot dummy2_t + \\
+\;& (\underset{(0.0722)}{-0.1739} - \underset{(0.0808)}{0.2378} \cdot v_{t-1} - \underset{(0.2795)}{0.5881} \cdot \Delta v_{t-4} + \underset{(0.0113)}{0.0400} \cdot gap_t^2) \cdot \\
\cdot\;& [1 - exp\{-\underset{(0.5963)}{1.2706}(\Delta v_{t-1} - \underset{(0.0383)}{0.1142})^2/0.3085\}]
\end{aligned} \tag{6.18}$$

The estimate of the coefficient c makes sense, because it lies in the range of the transition variable. The low value of $\gamma = 1.27$ indicates slow transition between the two extreme regimes. The variable gap^2 is significant in the nonlinear part, but insignificant in the linear part.

Finally, specification and diagnostic tests are performed to evaluate the obtained model. The p-values of the Jarque-Bera test and the test of no remaining error autocorrelation show that the null hypotheses of the normal distribution of the error term and of no error autocorrelation, respectively, cannot be rejected (Tables 6.22 and 6.23). One can see from Table 6.24 that there are no ARCH effects present in our model. The test of parameter constancy detects only problems concerning the constant term in the nonlinear part of the model (Table 6.25).

Jarque-Bera test p-value	0.6495
Ljung-Box p-value (8 lags)	0.8926
AIC	-3.4703
R^2	0.7275
S. E. of residuals	0.1686
Ratio var(nonlin)/var(lin)	0.6540
No. of observations	127

Table 6.22: Specification and diagnostic tests

No. of Lags	F-statistics	Deg. of freedom	p-value
1	1.3672	1, 114	0.2447
2	0.7417	2, 113	0.4786
3	1.2626	3, 112	0.2907
4	0.9568	4, 111	0.4343
5	0.8579	5, 110	0.5119
6	0.7409	6, 109	0.6179

Table 6.23: LM test of no remaining error autocorrelation

Lag	Autocorrelation	McLeod-Li p-value	ARCH LM p-value
1	0.0960	0.2203	0.2290
2	0.0314	0.4629	0.4527
3	0.1038	0.3122	0.2843
4	-0.0108	0.4750	0.4107
5	0.0681	0.6266	0.5323
6	-0.0272	0.5358	0.5796

Table 6.24: McLeod-Li and ARCH tests

Test	F-statistic	Deg. of freedom	p-value
LM3	5.1279	3, 111	0.0023
LM2	6.4165	2, 112	0.0023
LM1	12.4408	1, 113	0.0006

Table 6.25: Parameter constancy tests for the constant term in the nonlinear part of the model

By plotting the transition function against time, the frequent changes between the two extreme regimes become apparent (Figure 6.20). Both the value of 0 and the values near to 1 are attained several times during the sample period.

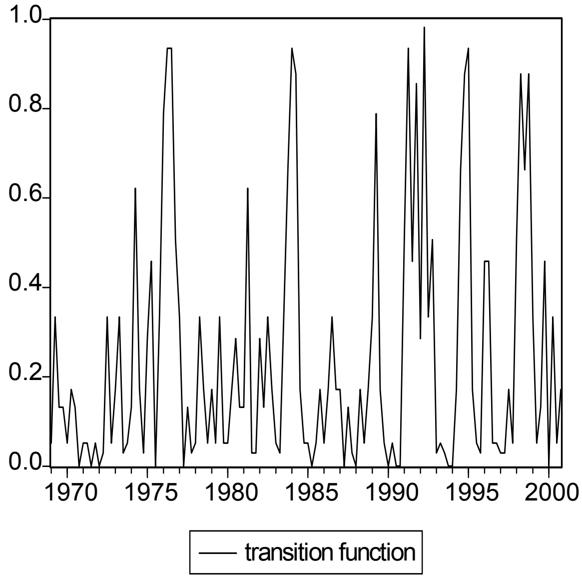

Figure 6.20: Transition function

A comparison of the linear and nonlinear model reveals an increase in explanatory power (R^2 increases from 0.63 to 0.73) and a decrease in the standard error of regression from 0.19 to 0.17. The null hypothesis of linearity tested against the alternative of a smooth transition regression model has to be rejected for every possible transition variable with the exception of the time trend. Both of these facts confirm our intuition that the linear relationship of Okun's law can be improved by consideration of regime changes.

By plotting the transition function G and the unemployment rate u in the same graph (Figure 6.21) one can observe that most of the major changes of the transition function occur when the unemployment rate has risen to new heights. This can be associated with major structural shifts in the German economy in those periods. As expected, one of the structural changes is detected in the early nineties, when the German reunion was taking place. Thus

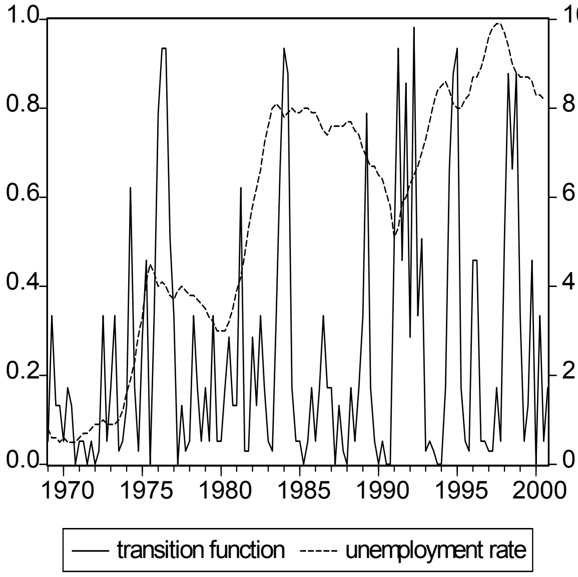

Figure 6.21: Unemployment rate and transition function

the nonlinear part of the model seems to capture the uneven development in the economy rather well.

Chapter 7

Conclusion

An important problem in the identification process of economic systems is usually related to the question if the model can be kept linear or whether nonlinear features are so dominant that they must be considered in the specification. From recent studies of univariate models one has learned that there is much to be gained by allowing a nonlinear specification. Representations of asymmetric reactions, structural changes and other phenomena of economic development can be fruitfully investigated by nonlinear modeling techniques. The smooth transition regression approach is one of the nonlinear techniques gaining importance in the econometric model building. The single equation smooth transition modeling has been studied extensively by several authors. Following the recent literature, chapters two and three summarize the various STR functional forms as well as the specification, estimation and evaluation procedure.

Since many issues in economics require the specification of several relationships, techniques to handle nonlinear features in systems are required. Only during the recent years such methods have appeared in the literature. As described in chapter four, Anderson and Vahid [2] devised a procedure for detecting common nonlinear components in a multivariate system of variables. The common nonlinearities approach is based on the canonical correlations technique and can help us interpret the relationships between different economic variables. The specification and estimation of the system of equations is also simplified, since the existence of common nonlinearities reduces the dimension of nonlinear components in the system and enables parsimony. This is particulary important in empirical investigations involving economic time series of shorter length. Namely, most of the macroeconomic indicators are published on a quarterly basis.

Weise [80], van Dijk [78] and Camacho [10] extended the univariate STR modeling approach developed by Teräsvirta and coworkers to vector autoregressive models of smooth transition. This approach at generalizing the single equation STR techniques is discussed in section 4.2. The smooth transition specification is limited to the case where the transition between different parameter regimes is governed by the same transition variable and the same type of transition function in every equation of the system. Weise, van Dijk and Camacho argue that since the economic practice imposes common nonlinear features, all equations share the same switching regime. But this argument is not convincing, since such a conclusion cannot be derived from economic theory, while applied econometric studies analyzing nonlinear systems are scarce. For this reason we extend their approach by allowing different smooth transition functional forms in different equations. The proposed extended specification procedure described in section 4.3 involves not only system linearity tests, but also single equation tests. Thus, systems including linear as well as nonlinear equations can be specified. All linearity tests are based on system estimates (to achieve efficiency) of a suitable auxiliary regression without restricting the choice of the transition variable. The decision rule for selecting the type of the transition function is also augmented to allow different types of the transition function in different equations of the system.

In chapter five, three-variable smooth transition vector autoregressive models of the consumer price index for Slovenia, consumer price index of another country and the nominal exchange rate between the currencies of both countries are discussed. The investigation thus applies the common nonlinearities techniques to small models of the real exchange rate, decomposed into its three components. The models for the five most important foreign trade partners of Slovenia, namely Germany, Italy, France, Austria and Croatia, are studied. The obtained real exchange rate model of the Slovenian Tolar versus Austrian Schilling, respectively Euro, contains only one common nonlinear component, as desired. On the other hand, models for Germany, Italy, France and Croatia cannot be adequately described with the help of only one type of nonlinearity. One of the possible explanations for such results could be the late accession of Austria to the European Union. Austria joined EU in the year 1995, whereas Germany, Italy and France were already member states in the year 1993, when we started our investigation. The process of Austria's EU accession had a deep impact on its economic structure and the relation to its neighbour states. In particular prices have been severely affected. These adjustments together with those ongoing in neighbouring Slovenia seem to be captured by a common nonlinear factor in the components of the real exchange rate. It concerns especially the effects of the lagged Austrian inflation rate and the nominal exchange rate, besides the cointegration terms. This lends economic support to the specification of a logistic smooth transition model with only one common nonlinear component.

A small nonlinear monetary model of inflation for West Germany is developed in chapter six. An equation describing the monetary system is augmented by the Phillips curve and the equation of Okun‘s law. It turns out that while the money demand equation and the Phillips curve exhibit nonlinear features, the Okun‘s law should be specified as a linear equation, since the single equation linearity tests cannot be rejected in any case. Additionally, the regime changes in the money demand equation and in the Phillips curve are governed by different transition variables. The estimated nonlinear system is inspected for its dynamic properties. Contrary to estimation, the money stock is considered a policy instrument to influence the inflationary process as well as the unemployment rate. The reaction of a change in oil prices, in the forward looking expected inflation rate and to a productivity shock is also evaluated. In the latter cases, reactions turn out to be rather symmetric, while shocks in the money stock exhibit significant asymmetries. This again shows that working with a nonlinear model generates a result that is known in practice, but not achieved by linear models.

The empirical investigations in chapters five and six again point out the drawbacks of the restricted smooth transition vector autoregressive specification proposed by Weise [80], van Dijk [78] and Camacho [10], imposing the same transition variable and the same type of the transition function in every equation. Chapter six reveals that the regime changes in the money demand equation and in the Phillips curve are governed by different transition variables, while only one of the five real exchange rate models studied in chapter five can be adequately described by one type of nonlinearity. Thus, the specification with the same transition variable and the same type of the transition function in every equation of a system is too restrictive and should not be imposed a priory.

Bibliography

[1] T. Amemiya: *Advanced Econometrics*, Harvard University Press, Cambridge, 1985.

[2] H.M. Anderson, F. Vahid: *Testing Multiple Equation Systems for Common Nonlinear Components*, Journal of Econometrics, **84**, p. 1–36, 1998.

[3] T.W. Anderson: *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1984.

[4] E. Berndt, B. Hall, R. Hall, J. Hausman: *Estimation and Inference in Nonlinear Structural Models*, Annals of Economic and Social Measurement, **3**, p. 653–665, 1974.

[5] O. Blanchard: *Macroeconomics*, Prentice Hall, New Jersey, 2000.

[6] B. Böhm: *Structural Change in the Austrian Inflation Process*, in G. Chaloupek, A. Guger, E. Nowotny (eds.) Ökonomie in Theorie und Praxis, p. 31 - 46, Springer Verlag, Berlin, 2001.

[7] B. Böhm: *Smooth Transition Models of Structural Change*, in J. Leskow, M. Puchet, L. F. Punzo (eds.) New Tools of Economic Dynamics, Lecture Notes in Economics and Mathematical Systems, Vol. **551**, Springer Verlag, New York, 2005.

[8] M. Burda, C. Wyplosz: *Macroeconomics: A European Text*, Oxford University Press, Oxford, 2005.

[9] M. Camacho, G. Perez – Quiros: *This is what the Leading Indicators Lead*, Journal of Applied Econometrics, **17**, p. 61 – 80, 2002.

[10] M. Camacho: *Vector Smooth Transition Regression Models for US GDP and the Composite Index of Leading Indicators*, Journal of Forecasting, **23**, p. 173-196, 2004.

[11] S.L. Chen, J.L. Wu: *Long-run Money Demand Revisited: Evidence from a Non-linear Approach*, Journal of International Money and Finance, **24**, p. 19–37, 2005.

[12] C. Chiarella, P. Flaschel, G. Gong, W. Semmler: *Nonlinear Phillips Curves, Complex Dynamics and Monetary Policy in a Keynesian Macro Model*, Chaos, Solitions and Fractals, **18**, p. 613–634, 2003.

[13] F. Collard, M. Juillard: *A Higher-Order Taylor Expansion Approach to Simulation of Stochastic Forward-Looking Models with an Application to a Nonlinear Phillips Curve Model*, Computational Economics, **17**, p. 125–139, 2001.

[14] L. Corrado, S. Holly: *Nonlinear Phillips Curves, Mixing Feedback Rules and the Distribution of Inflation and Output*, Journal of Economic Dynamics and Control, **28**, p. 467–492, 2003.

[15] J.C. Cuaresma: *Okun's Law Revisited*, Oxford Bulletin of Economics and Statistics, **65**, p. 439–451, 2003.

[16] R. Davidson, J. MacKinnon: *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993.

[17] J.J. Dolado, M.D. Ramon, M. Naveira: *Are Monetary-Policy Reaction Functions Asymmetric? The Role of Nonlinearity in the Phillips Curve*, European Economic Review, **49**, p. 485–503, 2005.

[18] Ø. Eitrheim, T. Teräsvirta: *Testing the Adequacy of Smooth Transition Autoregressive Models*, Journal of Econometrics, **74**, p. 59–76, 1996.

[19] R.F. Engle: *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the UK Inflation*, Econometrica, **50**, p. 987–1008, 1982.

[20] A. Escribano, O. Jordá: *Improved Testing and Specification of Smooth Transition Regression Models*, University of California Davis, Working Paper of Department of Economics 97–26, 1997.

[21] A. Escribano, S. Mira: *Nonlinear Time Series Models: Consistency and Asymptotic Normality of NLS under New Conditions*, Universidad Carlos III de Madrid, Statistics and Econometrics Series 14, Working Paper 95–42, 1995.

[22] R. Fletcher: *Practical Methods of Optimization*, Wiley, New York, 1987.

[23] H. Frisch: *Theories of Inflation*, Cambridge University Press, Cambridge, 1983.

[24] A.B.C. Galvao: *Multivariate Threshold Models: TVARs and TVECMs*, Brasilian Review of Econometrics, **23**, p. 143–171, 2003.

[25] L.G. Godfrey: *Misspecification Tests in Econometrics*, Cambridge University Press, Cambridge, 1988.

[26] R. Gordon: *The Time - varying NAIRU and its Implications for the Economic Policy*, Journal of Economic Perspectives, **11**, p. 11 - 32, 1997.

[27] C.W.J. Granger, T. Teräsvirta: *Modelling Nonlinear Economic Relationships*, Oxford University Press, New York, 1993.

[28] W.H. Greene: *Econometric Analysis*, Prentice Hall, New Jersey, 2003.

[29] R.E. Hall, J.B. Taylor; *Macroeconomics*, W. W. Norton & Company, New York, 1993.

[30] E.M. Hamann: *Time Series Models for the West German Economy*, Vienna Technical University, Vienna, 2001.

[31] B.E. Hansen: *Lagrange Multiplier Tests for Parameter Instability in Non-linear Models*, Paper presented at the Sixth World Congress of the Econometric Society, Barcelona, 1990.

[32] A.C. Harvey: *The Econometric Analysis of Time Series*, LSE Handbooks in Economics, Philip Allan, London, 1990.

[33] F. Hayashi: *Econometrics*, Princeton University Press, New Jersey, 2000.

[34] J.B. Hill: *Consistent LM - Tests for Linearity Against Compound Smooth Transition Alternatives*, Econometric Society 2004 North American Summer Meetings, **42**, 2004.

[35] J.B. Hill: *LM - Tests for Linearity Against Compound Smooth Transition Alternatives: A Bootstrap Simulation Study*, Economics Working Paper Archive at WUSTL, Series Econometrics, 2004.

[36] M.A. Hooker: *Are Oil Shocks Inflationary? Asymmetric and Nonlinear Specifications versus Changes in Regime*, Journal of Money, Credit, and Banking, **34**, p. 540–561, 2002.

[37] H.S. Huh: *Estimating Asymmetric Output Cost of Lowering Inflation for Australia*, Southern Economic Journal, **68**, p. 600–616, 2002.

[38] E.S. Jansen, T. Teräsvirta: *Testing Parameter Constancy and Super Exogeneity in Econometric Equations*, Oxford Bulletin of Economics and Statistics, **58**, p. 735–763, 1996.

[39] J. Johnston, J. DiNardo: *Econometric Methods*, Mc - Graw Hill, Singapore, 1997.

[40] A. Kavkler, B. Böhm: *Modelling Okun's Law for Germany with Smooth Transition Regression*, Proceedings of the 7th International Symposium on Operational Research, Podčetrtek, Slovenia, p. 289–294, 2003.

[41] A. Kavkler, B. Böhm: *Using Canonical Correlations in Testing for Common Nonlinear Components*, Program and Abstracts of the International Conference Applied Statistics 2005, Bled, Slovenia, p. 58, 2005.

[42] A. Kavkler, B. Böhm: *A Nonlinear Monetary Model of Inflation*, Proceedings of the 8th International Symposium on Operational Research, Nova Gorica, Slovenia, p. 291–296, 2005.

[43] R.G. King, C.I. Plosser, S.T. Rebelo: *Production, Growth and Business Cycles II: New Directions*, Journal of Monetary Economics, **21**, p. 309–342, 1988.

[44] L.A. Klimko, P.I. Nelson: *On Conditional Least Squares Estimation for Stochastic Processes*, Annals of Statistics, **6**, p. 629–642, 1978.

[45] G.M. Koop, H. Pesaran, S.M. Potter: *Impulse Response Analysis in Nonlinear Multivariate Models*, Journal of Econometrics, **74**, p. 119–147, 1996.

[46] H.M. Krolzig: *Markov–Switching Vector Autoregressions: Modelling, Statistical Inference and Applications to Business Cycle Analysis*, Lecture Notes in Economics and Mathematics, **454**, Springer Verlag, Berlin, 1997.

[47] T.L. Lai, C.Z. Wei: *Least Squares Estimates in Stochastic Regression Models with Applications in Identification and Control of Dynamic Systems*, Annals of Statistics, **10**, p. 154–166, 1982.

[48] C.F. Lin, T. Teräsvirta: *Testing the Constancy of Regression Parameters against Continuous Structural Change*, Journal of Econometrics, **62**, p. 211–228, 1994.

[49] H. Lütkepohl, T. Teräsvirta, J. Wolters: *Investigating Stability and Linearity of a German M1 Money Demand Function*, Journal of Applied Econometrics, **14**, p. 511–525, 1999.

[50] R. Luukkonen, P. Saikkonen, T. Teräsvirta: *Testing Linearity against Smooth Transition Autoregressive Models*, Biometrika, **75**, p. 491–499, 1998.

[51] R. Luukkonen, P. Saikkonen, T. Teräsvirta: *Testing Linearity in Univariate Time Series*, Scandinavian Journal of Statistics, **15**, p. 161–175, 1998.

[52] D.S. Maddala: *Econometrics*, McGraw - Hill, New York, 1977.

[53] D.G. Mayes, M. Viren: *Asymmetry and the Problem of Aggregation in the Euro Area*, Empirica, **29**, p. 47–73, 2002.

[54] G.E. Mizon, D.F. Hendry: *An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification*, Review of Economic Studies, **47**, p. 21–45, 1980.

[55] A.I. McLeod, W.K. Li: *Diagnostic Checking ARMA Time Series Models Using Squared Residuals*, Journal of Time Series Analysis, **4**, p. 269–273, 1983.

[56] A.R. Nobay, D.A. Peel: *Optimal Monetary Policy with a Nonlinear Phillips Curve*, Economics Letters, **67**, p. 159–164, 2000.

[57] J. Nocedal, S. Wright: *Numerical Optimization*, Springer, New York, 1999.

[58] J. Nyblom: *Testing for the Constancy of Parameters over Time*, Journal of the American Statistical Association, **84**, p. 223–230, 1989.

[59] J. Nyblom, T. Mäkeläinen: *Comparison of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model*, Journal of the American Statistical Association, **78**, p. 856–864, 1983.

[60] P. Pasching: *Modellierung der Phillipskurve mittels Smooth Transition Regression Modellen*, Vienna Technical University, Vienna, 2001.

[61] P. Petrović, Z. Mladenović: *Money Demand and Exchange Rate Determination under Hyperinflation: Conceptual Issues and Evidence from Yugoslavia*, Journal of Money, Credit, and Banking, **32**, p. 785–806, 2000.

[62] A.W. Phillips: *The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957*, Economica, **22**, p. 283–299, 1958.

[63] M.F.J. Prachowny: *Okun's Law: Theoretical Foundations and Revised Estimates*, Review of Economics and Statistics, **75**, p. 331–336, 1993.

[64] P.A. Samuelson, R.M. Solow: *Analytical Aspects of Anti-Inflation Policy*, American Economic Review, **50**, p. 177-194, 1960.

[65] L. Sarno: *Adjustment Costs and Nonlinear Dynamics in the Demand for Money: Italy, 1861-1991*, International Journal of Finance and Economics, **4**, p. 155–177, 1999.

[66] L. Sarno, M.P. Taylor, D.A. Peel: *Nonlinear Equilibrium Correction in U.S. Real Money Balances, 1869-1997*, Journal of Money, Credit, and Banking, **35**, p. 787–799, 2003.

[67] E. Schaling: *The Nonlinear Phillips Curve and Inflation Forecast Targeting: Symmetric versus Asymmetric Monetary Policy Rules*, Journal of Money, Credit, and Banking, **36**, p. 361–386, 2004.

[68] L. Sögner, A. Stiassny: *An Analysis on the Structural Stability of the Okun's Law - a Cross-Country Study*, Applied Economics, **14**, p. 1775–1787, 2002.

[69] A. Spanos: *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge, 1986.

[70] E.W. Tallman, D.P. Tang, P. Wang: *Nominal and Real Disturbances and Money Demand in Chinese Hyperinflation*, Economic Inquiry, **41**, p. 234–249, 2003.

[71] D.N. Tambakis: *Monetary Policy with a Nonlinear Phillips Curve and Asymmetric Loss*, Studies in Nonlinear Dynamics and Econometrics, **3**, p. 223–237, 1999.

[72] T. Teräsvirta: *Modelling Economic Relationships with Smooth Transition Regression*, Handbook of Applied Economic Statistics, p. 507–552, Marcel Dekker, New York, 1998.

[73] T. Teräsvirta: *Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models*, Journal of the American Statistical Association, **89**, p. 208–218, 1994.

[74] T. Teräsvirta, H.M. Anderson: *Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models*, Journal of Applied Econometrics, **7**, p. 119–136, 1992.

[75] T. Teräsvirta, D. van Dijk, M. Medeiros: *Linear Models, Smooth Transition Autoregressions and Neural Networks for Forecasting Macroeconomic Time Series: A Re-examination*, International Journal of Forecasting, **21**, p. 755–774, 2005.

[76] H. Tong: *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford, 1990.

[77] R.S. Tsay: *Testing and Modeling Multivariate Threshold Models*, Journal of the American Statistical Association, **93**, p. 1188–1202, 1998.

[78] D. van Dijk: *Smooth Transition Models: Extensions and Outlier Robust Inference*, Tinberg Institute, Amsterdam, 1999.

[79] C.E. Weber, J.E. West: *Functional Form in Regression Models of Okun's Law*, Applied Economics Letters, **3**, p. 607–609, 1996.

[80] C.L. Weise: *The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach*, Money, Credit and Banking, **31**, p. 85–108, 1999.

[81] H. White: *Asymptotic Theory for Econometricians*, Academic Press, San Diego, 2001.

[82] J. Wolters, T. Teräsvirta, H. Lütkepohl: *Modeling the Demand for M3 in the Unified Germany*, The Review of Economics and Statistics, **80**, p. 399–409, 1998.

[83] J.M. Wooldridge: *Estimation and Inference for Dependent Processes*, Handbook of Econometrics, Vol. **4**, Chapter 45, p. 2649–2738, North - Holland, Amsterdam, 1994.

Appendix A

Testing procedures

Because of the importance of the asymptotic distributions of the test statistics from the previous chapters - especially the Lagrange multiplier test - for the smooth transition regression models, a special chapter is devoted to this topic.

A.1 Maximum likelihood estimation

A comprehensive discussion about the method of maximum likelihood can be found in Harvey [32] and Greene [28]. The summary in this and the next sections of this chapter is based on both books, whereas the notation is the same as in Greene and the asymptotic theory theorems are taken from Amemiya [1], White [81] and Davidson and MacKinnon [16].

Suppose that we draw a random sample $y = (y_1, y_2, \ldots, y_n)'$ with n elements from a distribution with the probability density function $f(x|\theta)$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$ is a vector of parameters of the given distribution. The parameter vector θ is often not known and has to be estimated. The joint density of the obtained independent and identically distributed observations is of the form

$$f(y_1, y_2, \ldots, y_n|\theta) = \prod_{i=1}^{n} f(y_i|\theta). \tag{A.1}$$

The function $f(y_1, y_2, \ldots, y_n|\theta)$ may be interpreted in another way. For a fixed random sample $y = (y_1, y_2, \ldots, y_n)'$, $f(y_1, y_2, \ldots, y_n|\theta)$ depends only on the parameter vector θ, therefore we can define a new function as

$$L(\theta|y_1, y_2, \ldots, y_n) = f(y_1, y_2, \ldots, y_n|\theta). \tag{A.2}$$

Definition A.1. *The function $L(\theta|y)$ is called the likelihood function.*

L is a function of the parameter vector θ, conditioned on the data. In case of a discrete random vector y, $L(\theta|y)$ denotes the probability of drawing the random sample y that has actually been drawn. The interpretation is not the same for a continuous random vector y, since the probability of drawing any particular sample is zero, but the principle is similar.

Definition A.2. *The term maximum likelihood estimator (MLE) is often used to mean two different concepts:*

(1) *the value of θ that globally maximizes the likelihood function over the parameter space Θ,*

(2) *any root of the likelihood equation*

$$\frac{\partial L(\theta|y)}{\partial \theta} = 0. \tag{A.3}$$

We will use the definition (2) and refer to the definition (1) as the *global maximum likelihood estimator*. Note that a solution of the likelihood equation fulfills only the necessary condition for a local extremum. If, in addition, the Hessian matrix of second derivatives is negative definite, then the solution corresponds to a local maximum of the likelihood function.

Instead of the likelihood function L, one usually uses the *log - likelihood function* l, defined as

$$l(\theta|y) = \ln L(\theta|y) = \sum_{i=1}^{n} \ln f(y_i|\theta), \tag{A.4}$$

which is easier to work with. Our next task is to state the large - sample properties of the maximum likelihood estimator $\hat{\theta}$. Let us first explain the so-called regularity conditions, under which one is able to derive the asymptotic properties. Note that the derivations are computed with respect to θ, while y_i, $i = 1, 2, \ldots, n$, are treated as random variables in conditions R2 and R3.

Definition A.3 ([16], [28]). *The regularity conditions are:*

R1. *The function $\ln f(y_i|\theta)$ is at least three times continuously differentiable with respect to θ, for all y_i and for all $\theta \in \Theta$.*

R2. *The expectations of the first and the second derivatives of* $\ln f(y_i|\theta)$ *exist.*

R3. *There exists a function* z *such that for all* y_i

$$\left|\frac{\partial^3 \ln f(y_i|\theta)}{\partial\theta_j \partial\theta_k \partial\theta_l}\right| < z(y_i|\theta) \;\; \forall j,k,l \in \{1,2,\ldots,m\},\, \forall\theta \in \Theta, \tag{A.5}$$

and $z(y_i|\theta)$ *has a finite expectation.*

Let us denote the true value of the parameter vector by θ_0. The partial derivatives $\frac{\partial \ln f(y_i|\theta)}{\partial\theta}$ and $\frac{\partial^2 \ln f(y_i|\theta)}{\partial\theta\partial\theta'}$ are denoted by g_i and H_i, respectively, and the notation $g_i(\theta_0)$ is used to indicate the value of the function g_i evaluated at the true parameter vector θ_0. Hence, the *score vector*

$$g = \frac{\partial \ln L(\theta|y)}{\partial\theta} \tag{A.6}$$

is the sum of the vectors g_i, $i = 1,\ldots,n$, and the Hessian of the log-likelihood function $\ln L(\theta|y)$ is

$$H = \frac{\partial^2 \ln L(\theta|y)}{\partial\theta\partial\theta'} = \sum_{i=1}^{n} H_i. \tag{A.7}$$

The *information matrix*, which plays an important role in the derivation of the minimum variance bound and asymptotic efficiency of the maximum likelihood estimator, is defined as

$$I(\theta_0) = -E\left(\frac{\partial^2 \ln L(\theta|y)}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_0}\right). \tag{A.8}$$

For the class of unbiased estimators, the minimum variance bound is determined by the well-known Cramér-Rao theorem. Since the MLE is often biased, the minimum variance bound and asymptotic efficiency refer to the class of all consistent, asymptotically normally distributed estimators, abbreviated as CAN.

Definition A.4. *An estimator is asymptotically efficient, if it belongs to the class of CAN estimators and if its asymptotic covariance matrix is not larger than the asymptotic covariance matrix of any other CAN estimator (i.e. the asymptotic covariance matrix of the less efficient estimator equals that of the efficient estimator plus a positive semidefinite matrix).*

Definition A.5. *If* $\sqrt{n}(z_n - \mu) \xrightarrow{d} N(0, V)$*, then the random vector* z_n *is asymptotically normally distributed,* $z_n \overset{a}{\sim} N(\mu, \frac{1}{n}V)$*, and the asymptotic covariance matrix is denoted by* $AsyVar(z_n) = \frac{1}{n}V$*.*

The expectation and the variance of the score vector are given by the next proposition.

Proposition A.6. *If the regularity conditions R1 to R3 are fulfilled, then*

$$E\left(\frac{\partial \ln L(\theta|y)}{\partial \theta}\bigg|_{\theta=\theta_0}\right) = E(g(\theta_0)) = 0, \tag{A.9}$$

and

$$\begin{aligned} Var\left(\frac{\partial \ln L(\theta|y)}{\partial \theta}\bigg|_{\theta=\theta_0}\right) &= Var(g(\theta_0)) \\ &= E\left[\left(\frac{\partial \ln L(\theta|y)}{\partial \theta} \cdot \frac{\partial \ln L(\theta|y)}{\partial \theta'}\right)\bigg|_{\theta=\theta_0}\right] \\ &= I(\theta_0). \end{aligned} \tag{A.10}$$

The second result is known as the information matrix equality*.*

We can now state the large-sample properties of the maximum likelihood estimator.

Theorem A.7 ([28], Theorem 17.1 and Theorem 17.4). *Under the regularity conditions R1 - R3, the maximum likelihood estimator* $\hat{\theta}$ *has the following asymptotic properties:*

(i) *Consistency:* $plim\hat{\theta} = \theta_0$*.*

(ii) *Asymptotic normality:* $\hat{\theta} \overset{a}{\sim} N\left(\theta_0, I(\theta_0)^{-1}\right)$ *and* $AsyVar(\hat{\theta}) = I(\theta_0)^{-1}$*.*

(iii) *Asymptotic efficiency:* $\hat{\theta}$ *is asymptotically efficient.*

(iv) *Cramér-Rao lower bound: Minimum variance bound for a CAN estimator of the parameter vector* θ_0 *is given by*

$$I(\theta_0)^{-1} = \left(-E\left(\frac{\partial^2 \ln L(\theta|y)}{\partial \theta \partial \theta'}\bigg|_{\theta=\theta_0}\right)\right)^{-1}, \tag{A.11}$$

i.e. if $\hat{\psi}$ is any consistent, asymptotically normally distributed estimator of the parameter vector θ_0, then $AsyVar(\hat{\psi}) \geq I(\theta_0)^{-1}$. It follows from (i) and (ii) that the minimum variance bound is achieved by the MLE $\hat{\theta}$.

A sketch of the proof can be found in Greene [28] and an exact proof in Davidson and MacKinnon [16].

The asymptotic covariance matrix $I(\theta_0)^{-1} = \left(-E\Big(\frac{\partial^2 \ln L(\theta|y)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}\Big)\right)^{-1}$ contains the parameter vector θ_0, which has yet to be estimated. Hence, the question how to estimate the asymptotic covariance matrix arises. Greene [28] describes three possibilities:

(1) Under the assumption that the expected values of the second derivatives of the log-likelihood function with respect to θ are known, one may simply calculate their value at the maximum likelihood estimator $\hat{\theta}$. But in practice this is rarely the case.

(2) To avoid deriving the expected values, the second derivatives are evaluated at $\hat{\theta}$:

$$\hat{I}(\hat{\theta})^{-1} = \left(-\left(\frac{\partial^2 \ln L(\theta|y)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}}\right)\right)^{-1}. \tag{A.12}$$

The obtained estimator can be justified by observing that instead of the expected values the sample means were used. The law of large numbers and the consistency of $\hat{\theta}$ guarantee the consistency of this estimator.

(3) An even simpler estimator

$$\hat{\hat{I}}(\hat{\theta})^{-1} = \left(\sum_{i=1}^{n} \hat{g}_i \hat{g}_i'\right)^{-1} = \left(\hat{G}'\hat{G}\right)^{-1}, \tag{A.13}$$

with $\hat{g}_i = g_i(\hat{\theta})$ and $\hat{G} = \left(\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_n\right)'$, is based on the information matrix equality from Proposition A.6, which states that the information matrix equals the covariance matrix of the first derivative of the log-likelihood function. This estimator is called the *BHHH estimator* (since it was first advocated by Berndt, Hall, Hall and Hausman, see [4]) or the *outer product of gradients estimator* (for obvious reasons). The central limit theorem guarantees the consistency of the BHHH estimator (see Davidson and MacKinnon [16] for details).

A.2 Classical testing procedures

Suppose we would like to test if the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$ satisfies a set of q conditions or restrictions. If the restrictions are linear, they can be written as

$$R\theta = r, \tag{A.14}$$

where R is a full rank matrix with dimensions $q \times m$ and r is a vector of length q. To test such a hypothesis, three asymptotically equivalent test procedures are available. In a more general formulation, nonlinear restrictions are also allowed. In this case, the restrictions under the null hypothesis take the form

$$H_0 : c(\theta) = r, \tag{A.15}$$

where $c(\theta) = (c_1(\theta), c_2(\theta), \ldots, c_q(\theta))'$ and the functions $c_j(\theta)$, $j = 1, 2, \ldots, q$, are continuously differentiable. The alternative hypothesis is

$$H_1 : c(\theta) \neq r. \tag{A.16}$$

A comprehensive discussion about the derivation of the tests and their asymptotic distributions under the null hypothesis of nonlinear restrictions is given in White [81].

A.2.1 Likelihood ratio test

Let us denote by $\hat{\theta}_U$ the unrestricted maximum likelihood estimator and by $\hat{\theta}_R$ the restricted estimator, where the imposed restrictions are taken into account. If $\hat{L}_U$ and $\hat{L}_R$ stand for the value of the likelihood function evaluated at the estimators $\hat{\theta}_U$ and $\hat{\theta}_R$, respectively, and if the restrictions hold, then intuitively $\hat{L}_R$ should be close to $\hat{L}_U$. In other words, the *likelihood ratio*, defined as

$$\lambda = \frac{\hat{L}_R}{\hat{L}_U}, \tag{A.17}$$

should be close to 1. λ cannot be greater than 1, since the restricted maximum cannot exceed the unrestricted one. The *likelihood ratio test statistic* is of the form

$$LR = -2 \ln \lambda. \tag{A.18}$$

As we shall see later, the logarithm of the likelihood ratio is used in the test statistic in order to obtain a known distribution.

The LR statistic suffers the disadvantage of requiring both the unrestricted and the restricted estimator. In spite of that, it is popular, since it is easy to calculate.

A.2.2 Wald test

The Wald test is based on the observation that when the null hypothesis

$$H_0 : c(\theta) - r = 0 \tag{A.19}$$

holds, $c(\hat{\theta}_U) - r$ should be close to zero. The *Wald test statistic* is

$$W = \Big(c(\hat{\theta}_U) - r\Big)' \Big(AsyVar\big(c(\hat{\theta}_U) - r\big)\Big)^{-1} \Big(c(\hat{\theta}_U) - r\Big). \tag{A.20}$$

The Wald test is used when the unrestricted estimator is easier to compute than the restricted estimator. To construct the Wald test, one has to derive the asymptotic covariance matrix of the (possibly) nonlinear vector function $c\big(\hat{\theta}_U\big)$. This can be done by using the next proposition.

Proposition A.8 ([28], Theorem D.22 and Theorem 6.1). *If $\hat{\theta}_n$ is an m-vector of consistent asymptotically normally distributed parameter estimators, i.e.*

$$\begin{aligned} plim\, \hat{\theta}_n &= \theta_0, \\ \hat{\theta}_n &\overset{a}{\sim} N\Big(\theta_0, V_n\Big), \end{aligned} \tag{A.21}$$

and if $c(\theta)$ is a vector of q continuously differentiable functions (with respect to θ) not involving n, then

$$c(\hat{\theta}_n) \overset{a}{\sim} N\Big(c(\theta_0), C(\theta_0)V_n C(\theta_0)'\Big), \tag{A.22}$$

where $C(\theta_0) = \frac{\partial c(\theta)}{\partial \theta'}|_{\theta=\theta_0}$ is a $q \times m$-matrix of partial derivatives.

The proof of the proposition is based on the Slutsky theorem, saying that for a continuous function g and for a sequence x_n converging in probability, $plim\, g(x_n)$ is equal to $g(plim\, x_n)$. From the Slutsky theorem it follows that

$$plim\, c(\hat{\theta}_n) = c(\theta_0). \tag{A.23}$$

Let us briefly describe the idea of the proof. Since $plim\,\hat{\theta}_n = \theta_0$, the higher-order terms in the Taylor expansion

$$c(\hat{\theta}_n) = c(\theta_0) + C(\theta_0)(\hat{\theta}_n - \theta_0) + higher - order\,terms \tag{A.24}$$

can be neglected in large samples, therefore $c(\hat{\theta}_n)$ has the same asymptotic distribution as $c(\theta_0) + C(\theta_0)(\hat{\theta}_n - \theta_0)$ and

$$AsyVar(c(\hat{\theta}_n)) = C(\theta_0)AsyVar(\hat{\theta}_n)C(\theta_0)', \tag{A.25}$$

as stated in the proposition.

The true parameter value θ_0 is not known, hence the MLE $\hat{\theta}_U$ is used in its place to obtain the estimated asymptotic covariance matrix:

$$\hat{C} = \frac{\partial c(\theta)}{\partial \theta'}\bigg|_{\theta=\hat{\theta}_U} \tag{A.26}$$

and

$$EstAsyVar\big(c(\hat{\theta}_U) - r\big) = \hat{C}\Big(EstAsyVar(\hat{\theta}_U)\Big)\hat{C}'. \tag{A.27}$$

In the special case of linear restrictions, when $c(\theta) = R\theta$, $\hat{C}$ is equal to R and

$$EstAsyVar\big(c(\hat{\theta}_U) - r\big) = R\Big(EstAsyVar(\hat{\theta}_U)\Big)R'. \tag{A.28}$$

From Theorem A.7 we know that the asymptotic covariance matrix of the MLE $\hat{\theta}$ is equal to $I(\theta_0)^{-1}$. We have already described 3 possibilities for estimating $I(\theta_0)^{-1}$ on page 146.

A.2.3 Lagrange multiplier test

The Lagrange multiplier test is used when the restricted estimator is easier to compute than the unrestricted estimator. A typical example is a nonlinear model, which becomes linear under the imposed restrictions. Intuitively, if the restrictions

$$c(\theta) - r = 0 \tag{A.29}$$

are valid, then the vector of partial derivatives $\frac{\partial \ln L(\theta)}{\partial \theta}$ (the score vector) evaluated at the restricted maximum likelihood estimator $\hat{\theta}_R$ will be close to zero. To derive the restricted MLE, let us define the Lagrangean function

$$\ln N(\theta) = \ln L(\theta) + \lambda'(c(\theta) - r), \tag{A.30}$$

where λ is the vector of Lagrange multipliers. After solving the system

$$\begin{aligned} \frac{\partial \ln N(\theta)}{\partial \theta} &= \frac{\partial \ln L(\theta)}{\partial \theta} + C'\lambda = 0, \\ \frac{\partial \ln N(\theta)}{\partial \lambda} &= c(\theta) - r = 0, \end{aligned} \tag{A.31}$$

with C equal to the derivatives matrix $\frac{\partial c(\theta)}{\partial \theta'}$, one obtains the restricted estimator $\hat{\theta}_R$ and the estimated vector of Lagrange multipliers $\hat{\lambda}$. As already mentioned,

$$\left.\frac{\partial \ln L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_R} = -\hat{C}'\hat{\lambda} = g(\hat{\theta}_R) \tag{A.32}$$

is supposed to be close to zero if restrictions (A.29) hold. Recall that by Proposition A.6 the covariance matrix of the score vector is equal to the information matrix. Thus, using similar reasoning as with the Wald test statistic, the Lagrange multiplier test statistic is defined by

$$LM = \left(\left.\frac{\partial \ln L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_R}\right)' \left(I(\hat{\theta}_R)\right)^{-1} \left(\left.\frac{\partial \ln L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_R}\right). \tag{A.33}$$

A.2.4 Asymptotic distribution of the three tests

As already mentioned, the three tests are asymptotically equivalent. To derive the asymptotic distribution of the tests, one has to consider the asymptotic normal distribution of the maximum likelihood estimator (see Theorem A.7) and the following proposition.

Proposition A.9 ([81], Corollary 4.28). *Let b_n be a random vector with k components and $b_n \overset{a}{\sim} N(0, V_n)$, with V_n a $k \times k$ positive definite matrix. Then*

$$b_n' V_n^{-1} b_n \overset{a}{\sim} \chi_k^2, \tag{A.34}$$

where χ_k^2 is a χ^2 random variable with k degrees of freedom.

Typically, the asymptotic covariance matrix V_n is unknown, but its consistent estimator in the form of the estimated asymptotic covariance matrix is often available. The use of the consistent estimator of the asymptotic covariance matrix is justified by the next proposition.

Proposition A.10 ([81], Theorem 4.30). *Let the assumptions of Proposition A.9 hold. If there exists a sequence of positive semidefinite symmetric matrices* $\{\hat{V}_n\}$ *such that* $V_n - \hat{V}_n \xrightarrow{P} 0$, *where* V_n *is O(1), and if there exists a real number* $\delta > 0$ *with the property that for all n sufficiently large* $det(V_n) > \delta > 0$, *then*

$$b_n' \hat{V}_n^{-1} b_n \overset{a}{\sim} \chi_k^2. \tag{A.35}$$

The asymptotic distributions of the three tests can now be stated.

Theorem A.11. *Provided that the regularity conditions R1 - R3 hold, the Likelihood ratio, Wald and Lagrange multiplier test statistics, defined by (A.18), (A.20) and (A.33), respectively, are asymptotically* χ_q^2 *- distributed under null hypothesis (A.19).*

For the proof, see Greene [28] or White [81]. It has to be emphasized that the small-sample properties of the three tests are usually not known.

Appendix B

Linear vector error correction models for the real exchange rate

This chapter describes preliminary specification of the linear vector error correction models (VECM) for the application of the common nonlinearities modeling approach. Three-variable models of domestic prices (P_t), foreign prices (P_t^*) and the nominal exchange rate between the currencies of the two countries (S_t) are constructed. The models have been applied to 5 most important foreign trade partners of Slovenia, namely Germany, Italy, France, Austria and Croatia. The econometric model employs variables expressed in growth rates with the help of the logarithmic transformation, therefore small letters are used to denote the transformed variables.

Firstly, unit root tests were applied to the variables p_t, p_t^* and s_t for each of the countries. Usually, augmented Dickey - Fuller (ADF) test with automatic lag length selection and maximal lag length of 12 was used. Intercept and time trend (denoted by ttrend) were included, when they were significant at the 5 % level. In case of doubt, other unit root tests, namely Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Phillips-Perron (PP) and and Ng-Perron (NP) were also conducted. It has to be emphasized that the null hypothesis is unit root in the series, with the exception of the Kwiatkowski-Phillips-Schmidt-Shin test, where the series is stationary under the null. All of the variables turned out to be integrated of order 1, or I(1). The results of the unit root tests for the variable p_t are given here, since this variable (logarithm of domestic prices, i.e. prices in Slovenia) is present in every model.

Variable: p_t		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept and trend)
p-value	0.1370	0.0000

Table B.1: Unit root test results for the variable p_t

The variable is integrated of order 1.

Next, cointegration tests were performed and the linear vector error correction models (VECM) were specified. The null hypothesis of no cointegrating relations could not be rejected only in case of Croatia, therefore a linear VAR model in the differenced variables Δp_t, Δp_t^* and Δs_t was specified. Orthogonal seasonal dummy variables, denoted by d1 to d12, were introduced into some of the models to reduce the autocorrelation effects. When performing the cointegration tests, the lag length was determined with the help of the 5 information criteria implemented in EViews (sequential modified LR test, final prediction error, Akaike information criterion, Schwarz information criterion and Hannan-Quinn information criterion). The lag length was set to the value that was optimal for the majority of the information criteria. The deterministic trend settings for the cointegration tests were specified according to the unit root test results of the 3 variables in each model.

B.1 Results for Germany

Results of the unit root tests (Table B.2): the variables p_t^* and s_t are both integrated of order 1. Cointegration test results in Table B.3 indicate 1 cointegrating equation of the form:

$$\begin{aligned} ce1_t &= p_{t-1} + \underset{(1.1559)}{0.5409} \cdot p_{t-1}^* - \underset{(0.2345)}{0.9757} \cdot s_{t-1} - \underset{(0.0013)}{0.0027} \cdot ttrend_t \\ &- 2.0702 \end{aligned} \tag{B.1}$$

Estimates of the linear vector error correction model can be found in Table B.4 and results of the common nonlinearities test in Table B.5.

UNIT ROOT TESTS FOR GERMANY		
Variable: p_t^*		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept)
p-value	0.0577	0.0000
Test	KPSS (level, intercept and trend)	
test statistic	$LM = 0.2845$	5 % critical value: 0.1460
Variable: s_t		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept)
p-value	0.0766	0.0001
Test	KPSS (level, intercept and trend)	
test statistic	$LM = 0.2117$	5 % critical value: 0.1460

Table B.2: Unit root test results for Germany

COINTEGRATION TEST RESULTS FOR GERMANY Lag length: 3		
Hypothesized number of CE(s)	Trace test statistic	5 % critical value
none	51.7855	42.44
at most 1	20.9089	25.32
at most 2	7.3941	12.25
Hypothesized number of CE(s)	Max-eigenvalue test statistic	5 % critical value
none	30.8766	25.54
at most 1	13.5148	18.96
at most 2	7.3941	12.25

Table B.3: Johansen cointegration test results for Germany

VECM(3) MODEL FOR GERMANY			
Sample: 1993:05 2003:12, Included observations: 128			
Variable	Δp_t equation	Δp_t^* equation	Δs_t equation
$ce1_t$	-0.0749 (0.0149)	0.0070 (0.0076)	-0.0353 (0.0169)
Δp_{t-1}	0.1684 (0.0869)	-0.0493 (0.0444)	0.1155 (0.0985)
Δp_{t-2}	0.0933 (0.0880)	0.0583 (0.0450)	-0.0867 (0.0997)
Δp_{t-3}	-0.0416 (0.0890)	0.1331 (0.0455)	-0.1911 (0.1008)
Δp_{t-1}^*	0.0156 (0.1814)	0.1026 (0.0928)	0.0608 (0.2056)
Δp_{t-2}^*	-0.1346 (0.1829)	-0.0464 (0.0935)	0.0847 (0.2073)
Δp_{t-3}^*	0.1559 (0.1743)	-0.2628 (0.0892)	0.0731 (0.1976)
Δs_{t-1}	0.1630 (0.0798)	0.0906 (0.0408)	0.7356 (0.0905)
Δs_{t-2}	-0.1877 (0.0996)	-0.0596 (0.0509)	-0.1512 (0.1129)
Δs_{t-3}	-0.0150 (0.0796)	0.0578 (0.0407)	-0.1814 (0.0902)
$const$	0.0061 (0.0011)	1.87E-05 (0.0006)	0.0038 (0.0012)
$d12_t$	-0.0018 (0.0016)	0.0013 (0.0008)	-0.0002 (0.0018)
R^2	0.4604	0.2492	0.5833
$S.E.$	0.0046	0.0023	0.0052
AIC	-7.8556	-9.1966	-7.6056
BIC	-7.5882	-8.9292	-7.3382
$AIC = -24.3918, \ BIC = -23.5006, \ \log L = 1619.976$			

Table B.4: Coefficient estimates (and p-values) of the VECM(3) model for Germany

COMMON NONLINEARITIES TEST FOR GERMANY									
tvar: $ce1_t$		tvar:$\Delta\ p_{t-1}$		tvar: Δp^*_{t-1}		tvar: Δs_{t-1}		tvar: $\Delta\ p_{t-2}$	
s	p-value	s	p-value	s	p-value	s	p-value	s	p-value
1	0.300	1	0.077	1	0.826	1	0.834	1	0.705
2	0.022	2	0.012	2	0.442	2	0.048	2	0.155
3	0.000	3	0.000	3	0.011	3	0.000	3	0.000
tvar: Δp^*_{t-2}		tvar: Δs_{t-2}		tvar: Δp_{t-3}		tvar:$\Delta\ p^*_{t-3}$		tvar: Δs_{t-3}	
s	p-value	s	p-value	s	p-value	s	p-value	s	p-value
1	0.610	1	0.685	1	0.851	1	0.290	1	0.998
2	0.320	2	0.030	2	0.244	2	0.070	2	0.460
3	0.152	3	0.000	3	0.000	3	0.070	3	0.000
df	s=1: 28, s=2: 58, s=3: 90								

Table B.5: Common nonlinearities test for Germany

B.2 Results for Italy

Results of the unit root tests (Table B.6): the variables p^*_t and s_t are both I(1). Trace test statistic from Table B.7 indicates 3 cointegrating equations, while max - eigenvalue test indicates 1 cointegrating equation of the form:

$$ce1_t = p_{t-1} - \underset{(0.6957)}{1.2874} \cdot p^*_{t-1} - \underset{(0.2512)}{1.1614} \cdot s_{t-1} + 7.2065 \qquad \text{(B.2)}$$

Estimates of the linear vector error correction model can be found in Table B.8 and results of the common nonlinearities test in Table B.9.

UNIT ROOT TESTS FOR ITALY		
Variable: p_t^*		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept and trend)
p-value	0.2913	0.0000
Variable: s_t		
Test	ADF (level, intercept)	ADF (1st diff., intercept)
p-value	0.1834	0.0000

Table B.6: Unit root test results for Italy

COINTEGRATION TEST RESULTS FOR ITALY Lag length: 1		
Hypothesized number of CE(s)	Trace test statistic	5 % critical value
none	53.0637	29.68
at most 1	22.6405	15.41
at most 2	9.4433	3.76
Hypothesized number of CE(s)	Max-eigenvalue test statistic	5 % critical value
none	30.4232	20.97
at most 1	13.1971	14.07
at most 2	9.4433	3.76

Table B.7: Johansen cointegration test results for Italy

VECM(1) MODEL FOR ITALY			
Sample: 1993:03 2003:12, Included observations: 130			
Variable	Δp equation	Δp^* equation	Δs equation
$ce1_t$	0.0090 (0.0037)	0.0059 (0.0013)	0.0576 (0.0196)
Δp_{t-1}	0.4298 (0.0850)	0.0105 (0.0289)	-0.3784 (0.4468)
Δp^*_{t-1}	0.2082 (0.2663)	-0.0408 (0.0895)	-1.4555 (1.4002)
Δs_{t-1}	0.0107 (0.0170)	0.0037 (0.0057)	0.2083 (0.0893)
$const$	0.0038 (0.0010)	0.0024 (0.0003)	0.0118 (0.0053)
$d1_t$	0.0107 (0.0017)	0.0018 (0.0006)	-8.54E-05 (0.0090)
$d2_t$	0.0043 (0.0018)	0.0016 (0.0006)	0.0138 (0.0093)
$d3_t$	0.0054 (0.0017)	0.0020 (0.0006)	-0.0023 (0.0087)
$d4_t$	0.0058 (0.0017)	0.0001 (0.0006)	-0.0039 (0.0088)
$d5_t$	0.0063 (0.0017)	0.0017 (0.0006)	0.0088 (0.0087)
$d7_t$	0.0056 (0.0017)	-0.0004 (0.0006)	-0.0034 (0.0087)
$d10_t$	0.0051 (0.0017)	0.0011 (0.0006)	-0.0035 (0.0087)
$d11_t$	0.0061 (0.0017)	0.0015 (0.0006)	0.0085 (0.0086)
$d9_t$	0.0095 (0.0017)	-0.0002 (0.0006)	0.0096 (0.0090)
$d12_t$	0.0039 (0.0016)	-0.0008 (0.0006)	0.0109 (0.0086)
R^2	0.5112	0.4452	0.1640
$S.E.$	0.0044	0.0015	0.0231
AIC	-7.9121	-10.093	-4.5925
BIC	-7.5812	-9.7622	-4.2616
$AIC = -22.2396, \quad BIC = -21.1808, \quad \log L = 1517.478$			

Table B.8: Coefficient estimates (and p-values) of the VECM(1) model for Italy

COMMON NONLINEARITIES TEST FOR ITALY							
tvar: $ce1_t$		tvar:Δ p_{t-1}		tvar: Δp^*_{t-1}		tvar: Δs_{t-1}	
s	p-value	s	p-value	s	p-value	s	p-value
1	0.789	1	0.488	1	0.230	1	0.176
2	0.255	2	0.164	2	0.033	2	0.038
3	0.000	3	0.008	3	0.000	3	0.000
df	s=1: 10, s=2: 22, s=3: 36						

Table B.9: Common nonlinearities test for Italy

B.3 Results for France

Results of the unit root tests (Table B.10): the variables p^*_t and s_t are both integrated of order 1. Cointegration test results in Table B.11 indicate 1 cointegrating equation of the form:

$$\begin{aligned} ce1_t &= p_{t-1} + \underset{(1.2230)}{2.4023} \cdot p^*_{t-1} - \underset{(0.3190)}{0.7141} \cdot s_{t-1} - \underset{(0.0017)}{0.0056} \cdot ttrend_t - \\ &- 11.8832 \end{aligned} \tag{B.3}$$

Estimates of the linear vector error correction model can be found in Table B.12 and results of the common nonlinearities test in Table B.13.

UNIT ROOT TESTS FOR FRANCE		
Variable: p_t^*		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept)
p-value	0.6511	0.0000
Variable: s_t		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept)
p-value	0.1202	0.0000
Test	KPSS (level, intercept and trend)	
test statistic	$LM = 0.5032$	5 % critical value: 0.1460

Table B.10: Unit root test results for France

COINTEGRATION TEST RESULTS FOR FRANCE Lag length: 3		
Hypothesized number of CE(s)	Trace test statistic	5 % critical value
none	52.0628	42.44
at most 1	21.3704	25.32
at most 2	2.9876	12.25
Hypothesized number of CE(s)	Max-eigenvalue test statistic	5 % critical value
none	30.6924	25.54
at most 1	18.3828	18.96
at most 2	2.9876	12.25

Table B.11: Johansen cointegration test results for France

VECM(3) MODEL FOR FRANCE			
Sample: 1993:05 2003:12, Included observations: 128			
Variable	Δp_t equation	Δp_t^* equation	Δs_t equation
$ce1_t$	-0.0471 (0.0086)	-0.0011 (0.0041)	-0.0547 (0.0200)
Δp_{t-1}	0.2636 (0.0950)	-0.0277 (0.0453)	-0.0425 (0.2204)
Δp_{t-2}	-0.0138 (0.0924)	0.0034 (0.0441)	-0.1294 (0.2145)
Δp_{t-3}	-0.0441 (0.0896)	0.0132 (0.0427)	-0.4689 (0.2079)
Δp_{t-1}^*	0.0636 (0.2086)	-0.0108 (0.0995)	-0.1814 (0.4841)
Δp_{t-2}^*	-0.1322 (0.2188)	-0.0924 (0.1044)	0.5511 (0.5078)
Δp_{t-3}^*	0.4835 (0.2114)	0.0413 (0.1008)	1.1207 (0.4907)
Δs_{t-1}	0.0227 (0.0392)	0.0081 (0.0187)	0.3419 (0.0910)
Δs_{t-2}	-0.1224 (0.0414)	-0.0037 (0.0198)	0.0929 (0.0961)
Δs_{t-3}	0.0244 (0.0406)	0.0075 (0.0194)	-0.2751 (0.0942)
$const$	0.0060 (0.0010)	0.0014 (0.0005)	0.0069 (0.0024)
$d1_t$	0.0107 (0.0015)	-7.57E-05 (0.0007)	0.0052 (0.0036)
$d2_t$	0.0057 (0.0016)	0.0024 (0.0008)	0.0028 (0.0040)
$d3_t$	0.0062 (0.0017)	0.0029 (0.0008)	-0.0002 (0.0040)
$d5_t$	0.0058 (0.0016)	0.0012 (0.0007)	-0.0022 (0.0036)
$d12_t$	0.0039 (0.0014)	0.0002 (0.0007)	-0.0015 (0.0033)
$d4_t$	0.0079 (0.0018)	0.0013 (0.0008)	0.0085 (0.0041)
$d7_t$	0.0048 (0.0015)	-0.0020 (0.0007)	0.0003 (0.0036)
$d9_t$	0.0077 (0.0018)	0.0019 (0.0009)	-0.0027 (0.0041)
$d10_t$	0.0063 (0.0017)	0.0008 (0.0008)	0.0075 (0.0040)
$d11_t$	0.0058 (0.0017)	-0.0001 (0.0008)	0.0024 (0.0038)
R^2	0.6629	0.3815	0.3649
$S.E.$	0.0038	0.0018	0.0087
AIC	-8.1854	-9.6662	-6.5016
BIC	-7.7175	-9.1983	-6.0337
$AIC = -23.9052,\ \ BIC = -22.4123,\ \ \log L = 1631.337$			

Table B.12: Coefficient estimates (and p-values) of the VECM(3) model for France

COMMON NONLINEARITIES TEST FOR FRANCE									
tvar: $ce1_t$		tvar:$\Delta\ p_{t-1}$		tvar: Δp^*_{t-1}		tvar: Δs_{t-1}		tvar: $\Delta\ p_{t-2}$	
s	p-value	s	p-value	s	p-value	s	p-value	s	p-value
1	0.413	1	0.662	1	0.681	1	0.648	1	0.121
2	0.020	2	0.067	2	0.649	2	0.001	2	0.041
3	0.000	3	0.000	3	0.150	3	0.000	3	0.000
tvar: Δp^*_{t-2}		tvar: Δs_{t-2}		tvar: Δp_{t-3}		tvar:$\Delta\ p^*_{t-3}$		tvar: Δs_{t-3}	
s	p-value	s	p-value	s	p-value	s	p-value	s	p-value
1	0.789	1	0.658	1	0.396	1	0.190	1	0.697
2	0.256	2	0.003	2	0.013	2	0.056	2	0.029
3	0.044	3	0.000	3	0.000	3	0.009	3	0.000
df	s=1: 28, s=2: 58, s=3: 90								

Table B.13: Common nonlinearities test for France

B.4 Results for Austria

Results of the unit root tests (Table B.14): the variables p^*_t and s_t are both integrated of order 1. Cointegration test results in Table B.15 indicate 2 cointegrating equations of the form:

$$\begin{aligned} ce1_t &= p^*_{t-1} - \underset{(0.6605)}{3.9639} \cdot s_{t-1} + \underset{(0.0030)}{0.0142} \cdot ttrend_t + 15.0858 \qquad \text{(B.4)} \\ ce2_t &= p_{t-1} + \underset{(0.3569)}{0.4985} \cdot s_{t-1} - \underset{(0.0016)}{0.0079} \cdot ttrend_t - 6.9070 \end{aligned}$$

Estimates of the linear vector error correction model can be found in Table B.16 and results of the common nonlinearities test in Table B.17.

UNIT ROOT TESTS FOR AUSTRIA		
Variable: p_t^*		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept)
p-value	0.1714	0.0000
Variable: s_t		
Test	ADF (level, intercept and trend)	ADF (1st diff., intercept)
p-value	0.0768	0.0001
Test	KPSS (level, intercept and trend)	
test statistic	$LM = 0.2117$	5 % critical value: 0.1460

Table B.14: Unit root test results for Austria

COINTEGRATION TEST RESULTS FOR AUSTRIA Lag length: 1		
Hypothesized number of CE(s)	Trace test statistic	5 % critical value
none	84.7897	42.44
at most 1	36.3163	25.32
at most 2	7.1575	12.25
Hypothesized number of CE(s)	Max-eigenvalue test statistic	5 % critical value
none	48.4734	25.54
at most 1	29.1588	18.96
at most 2	7.1575	12.25

Table B.15: Johansen cointegration test results for Austria

VECM(1) MODEL FOR AUSTRIA			
Sample: 1993:03 2003:12, Included observations: 130			
Variable	Δp_t equation	Δp_t^* equation	Δs_t equation
$ce1_t$	-0.0241 (0.0056)	-0.0043 (0.0040)	0.0360 (0.0074)
$ce2_t$	-0.0651 (0.0113)	-0.0073 (0.0080)	0.0299 (0.0148)
Δp_{t-1}	0.2276 (0.0831)	-0.0212 (0.0588)	0.1534 (0.1091)
Δp_{t-1}^*	-0.2740 (0.1291)	0.0870 (0.0914)	-0.4474 (0.1696)
Δs_{t-1}	0.0056 (0.0489)	0.0621 (0.0346)	0.6138 (0.0642)
$const$	0.0064 (0.0008)	0.0012 (0.0006)	0.0015 (0.0011)
$d1_t$	0.0095 (0.0015)	0.0006 (0.0011)	0.0011 (0.0020)
$d2_t$	0.0051 (0.0016)	0.0015 (0.0011)	-0.0010 (0.0020)
$d3_t$	0.0060 (0.0015)	0.0007 (0.0010)	0.0007 (0.0019)
$d4_t$	0.0062 (0.0014)	-0.0009 (0.0010)	0.0014 (0.0019)
$d5_t$	0.0060 (0.0015)	-0.0003 (0.0010)	-0.0021 (0.0019)
$d7_t$	0.0043 (0.0015)	0.0010 (0.0011)	-0.0023 (0.0020)
$d9_t$	0.0075 (0.0015)	-0.0030 (0.0011)	0.0010 (0.0020)
$d10_t$	0.0040 (0.0015)	-0.0005 (0.0011)	0.0007 (0.0020)
$d11_t$	0.0055 (0.0015)	-0.0014 (0.0010)	0.0025 (0.0019)
$d12_t$	0.0032 (0.0015)	4.99E-05 (0.0011)	-0.0005 (0.0019)
R^2	0.6180	0.2230	0.6692
$S.E.$	0.0039	0.0028	0.0051
AIC	-8.1433	-8.8335	-7.5975
BIC	-7.7904	-8.4805	-7.2445
$AIC = -24.1235, \quad BIC = -22.8882, \quad \log L = 1649.636$			

Table B.16: Coefficient estimates (and p-values) of the VECM(1) model for Austria

COMMON NONLINEARITIES TEST FOR AUSTRIA									
tvar: $ce1_t$		tvar: $ce2_t$		tvar: $\Delta\ p_{t-1}$		tvar: Δp^*_{t-1}		tvar: Δs_{t-1}	
s	p-value	s	p-value	s	p-value	s	p-value	s	p-value
1	0.214	1	0.262	1	0.172	1	0.286	1	0.490
2	0.055	2	0.079	2	0.010	2	0.024	2	0.132
3	0.001	3	0.001	3	0.000	3	0.000	3	0.000
df	s=1: 13, s=2: 28, s=3: 45								

Table B.17: Common nonlinearities test for Austria

UNIT ROOT TESTS FOR CROATIA		
Variable: p^*_t		
Test	ADF (level, intercept)	ADF (1st diff., intercept)
p-value	0.1792	0.0000
Variable: s_t		
Test	ADF (level, intercept)	ADF (1st diff., intercept)
p-value	0.6181	0.0000

Table B.18: Unit root test results for Croatia

B.5 Results for Croatia

Results of the unit root tests (Table B.18): the variables p^*_t and s_t are both integrated of order 1. Cointegration test results in Table B.19 indicate no cointegrating equations. Estimates of the linear vector autoregressive model in the differenced variables can be found in Table B.20 and results of the common nonlinearities test in Table B.21.

COINTEGRATION TEST RESULTS FOR CROATIA Lag length: 1		
Hypothesized number of CE(s)	Trace test statistic	5 % critical value
none	21.7833	29.68
at most 1	7.0661	15.41
at most 2	1.8141	3.76
Hypothesized number of CE(s)	Max-eigenvalue test statistic	5 % critical value
none	14.7172	20.97
at most 1	5.2520	14.07
at most 2	1.8141	3.76

Table B.19: Johansen cointegration test results for Croatia

VAR(1) MODEL FOR CROATIA Sample: 1995:06 2003:12, Included observations: 103			
Variable	Δp_t equation	Δp_t^* equation	Δs_t equation
Δp_{t-1}	0.1606 (0.0948)	0.0311 (0.1561)	0.2413 (0.2308)
Δp_{t-1}^*	0.0841 (0.0711)	0.2770 (0.1171)	0.0201 (0.1731)
Δs_{t-1}	0.1062 (0.0405)	0.0674 (0.0666)	0.3240 (0.0986)
$const$	0.0046 (0.0006)	0.0018 (0.0010)	0.0005 (0.0015)
$d6_t$	-0.0068 (0.0014)	-0.0090 (0.0023)	-0.0016 (0.0033)
$d8_t$	-0.0064 (0.0015)	-0.0052 (0.0025)	-0.0029 (0.0036)
$d12_t$	-0.0022 (0.0013)	0.0017 (0.0022)	-2.14E-05 (0.0032)
R^2	0.3821	0.2580	0.1324
$S.E.$	0.0037	0.0061	0.0091
AIC	-8.2862	-7.2889	-6.5064
BIC	-8.1072	-7.1099	-6.3274
$AIC = -21.9788, \ BIC = -21.4416, \ \log L = 1152.906$			

Table B.20: Coefficient estimates (and p-values) of the VAR(1) model for Croatia

COM. NONLIN. TEST FOR CROATIA					
tvar: $\Delta\ p_{t-1}$		tvar: Δp^*_{t-1}		tvar: Δs_{t-1}	
s	p-value	s	p-value	s	p-value
1	0.797	1	0.297	1	0.723
2	0.763	2	0.177	2	0.499
3	0.298	3	0.001	3	0.000
df	s=1: 7, s=2: 16, s=3: 27				

Table B.21: Common nonlinearities test for Croatia

Appendix C

Data and program sources

All the data used in the empirical investigations can be found on the enclosed CD. Here is a short description of the included files:

- The file *real exchange rate model.xls* contains the CPI time series for Slovenia, Austria, Germany, Italy, France and Croatia and the corresponding nominal exchange rate series used in chapter five. The data were obtained from the Bank of Slovenia and from the Statistical Office of the Republic of Slovenia.

- The data for the empirical investigation in chapter six were obtained from an OECD data disc and are given in the file *monetary model of inflation.xls*. The following time series are included: unemployment rate, inflation rate, forward looking price expectation, energy price inflation, labor productivity growth, real money growth and real growth rate.

- The data of M. Camacho used in section 4.3 can be downloaded from his web page:
  http://www.um.es/econometria/Maximo.
  The US composite index of leading indicators series and the US GDP series are given in files *cli.data* and *gdp.data*, respectively.

The specified models were estimated and evaluated with the help of the EViews and Gauss program packages. Several Gauss routines implementing the univariate STR modeling process were made available by professor

Teräsvirta and integrated into a user-friendly program by students of the Institute of mathematical methods in economics (previously Institute of econometrics, operations research and system theory). The common nonlinearities tests in chapter five were performed with the help of the Gauss programs written by professor Vahid that were adapted to suit our needs. For the calculations in section 4.3, the code written by Camacho (obtained from his web page) was modified to allow different transition variables in different equations of the system. Programs for single equation linearity tests (*first equation linearity test.gau* and *second equation linearity test.gau*) were developed with the help of the system linearity test code (*system linearity test.gau*).

Curriculum vitae

PERSONAL DATA:

Full name: Alenka Kavkler
Current address: Vrbanska cesta 12 b, 2000 Maribor, Slovenia
Date of birth: October 1st, 1971
Place of birth: Maribor, Slovenia
Citizenship: Slovene
Marital status: single

EDUCATION:

1986 - 1990: Secondary school with concentration in natural science, Maribor
Awards: 3rd prize on State Mathematics Competition in first and fourth grade

1990 - 1996: Bachelor of Science in Mathematics at the Faculty of Mathematics and Physics of the University of Ljubljana, Slovenia
Graduated with grade 10 (out of 10) and average grade of 9.55
Received Prešeren's prize for graduation thesis
Supervisor: Prof. Matjaž Omladič

1996 - 1999: Master of Science in Mathematics at the Faculty of Mathematics and Physics of the University of Ljubljana, Slovenia
Graduated with grade 10 and average grade of 9.75
Supervisor: Prof. Matjaž Omladič

2001 - 2006: Ph. D. student of Technical Mathematics with Concentration in Mathematical Methods in Economics at the Vienna University of Technology, Austria
Supervisor: Prof. Bernhard Böhm

WORK EXPERIENCE:

1996 - 1997: Faculty of Computer and Information Science, University of Ljubljana
Teaching assistant for mathematics

1997 - 2006: Faculty of Economics and Business, University of Maribor
Teaching assistant for the courses Mathematics in Economics and Mathematical Statistics