Nonlinear Econometric Methods in International Economics

Efthymios Pavlidis

BSc University of Patras, MSc Warwick University

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the Department of Economics, Lancaster University.

July 2009

Declaration

I hereby declare that this thesis is my own work and that it has not been submitted for any other degree.

Efthymios Pavlidis

Signature:

To Prof David A. Peel, Dr Ivan Paya, and my family for their support and motivation.

Abstract

This thesis builds upon recent developments in the areas of international economics, econometrics and computational statistics, to provide a robust framework for specifying, modelling and forecasting real exchange rates. The main research topics addressed are the following. First, the impact of conditional heteroskedasticity on linearity tests. Second, the parsimonious modelling and forecasting of the dollar-sterling real exchange rate using a long span of data. Third, the re-examination of the well-documented real exchange rate-consumption anomaly from the viewpoint of nonlinear dynamics. Finally, the relationship between real exchange rate persistence and time-varying trade costs.

Acknowledgements

I am grateful to my PhD supervisors, Prof David Peel and Dr Ivan Paya, for their exceptional guidance, support and, most of all, for stimulating my interest in applied economics. Their enthusiasm and imagination have been a source of inspiration throughout the last four years.

I am also indebted to my family for their everlasting encouragement and support. My brother, Nicos, deserves special thanks for teaching me how to do research in an efficient manner and for showing me the way the numerous times I got lost. For all the above and many more that I have not mentioned, this thesis is devoted to my supervisors and my family.

I thank Prof Costas Syriopoulos, Prof Michael Vrahatis, and Dr Dimitris Tasoulis for their valuable comments and insights. Finally, I express my gratitude to my colleagues in the Department of Economics for providing an excellent research environment and the administrator of the High Performance Cluster, Mike Pacey, for his excellent work and assistance.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning.

— Albert Einstein (1879 – 1955)

Over the last decades there has been a steadily increasing interest in the development of nonlinear time series models, and their application in international economics. This thesis focuses on a specific family of these models, the smooth transition autoregressive, and their usage in explaining and forecasting the behaviour of real exchange rates.

A natural starting point for the analysis of real exchange rates is the Purchasing Power Parity (PPP) theory. PPP states that the nominal exchange rate between two currencies should be equal to the ratio of aggregate price levels between the two countries, so that a unit of currency of one country will have the same purchasing power in a foreign country.[1] The hypothesis that PPP holds in the long run is

[1] Excellent surveys covering the origins of PPP theory and the findings of the associated empirical literature are provided in Sarno and Taylor (2002) and Taylor (2006).

a building block of many macroeconomic models. It is therefore of interest to international economists.

However, the first empirical studies employing unit root tests in the late 1980s were consistent in their failure to reject the unit root hypothesis for major real exchange rates (e.g., Taylor, 1988; Mark, 1990). Subsequent studies using longer time series data sets or panel methods suggested that the early non-rejections of the unit root hypothesis was due to low power of the corresponding test (Lothian and Taylor, 1996). Despite the evidence of mean reversion, the implied speeds of adjustment of the real exchange rate in these studies was implausibly slow, typically with half-life in the range of three to five years. Rogoff (1996) summarised this position as follows

> *"How can one reconcile the enormous short-term volatility of real exchange rates with the extremely slow rate at which shocks appear to damp out?"*
>
> Rogoff (1996, p. 647)

After the work of Rogoff (1996), perhaps the major change in emphasis has been the application of nonlinear rather than linear methods. These nonlinear models are based on theoretical analyses that embody factors such as transactions costs, limits to arbitrage and heterogeneity of expectations of market participants (see, e.g., Dumas, 1992; De Grauwe et al., 1993; Shleifer and Vishny, 1997). As a consequence, the real exchange rate is described by a nonlinear data generating process that exhibits a region of unit root (or near-unit root behaviour) near the equilibrium real exchange rate. Nonlinear models that capture this type of behaviour are the threshold autoregressive model of Tong (1983), and the exponential smooth transition autoregressive model of Ozaki (1978) and Teräsvirta (1994).

It follows that econometric modelling requires appropriate tests for linearity. Typically, researchers employ tests which are based on the assumption of homoskedastic residuals. However, the fact that changes in regime (e.g., fixed to floating, or different monetary regimes) may induce time-varying volatility raises concerns regarding statistical inferences. A number of authors have noted that the presence of conditional heteroskedasticity may lead to poor performance (spurious inference) of linearity tests (Lundbergh and Teräsvirta, 1998). Chapter 2 examines the robustness of conventional linearity tests and tests based on Bootstrap methods to conditional heteroskedasticity of unknown form. The importance of robust inference is highlighted through Monte Carlo simulations, as well as, several empirical applications on economic and financial time series data. The insights gained are, in turn, exploited in the remaining chapters.

Chapter 3 deals with modelling and forecasting the dollar-sterling real exchange rate using a long span of data. The motivation of the chapter is twofold. First, the empirical literature on the out-of-sample performance of nonlinear real exchange rate models is scarce. Second, there is a documented difficulty of nonlinear models to outperform their linear counterparts (Clements and Smith, 1999, see, e.g.,). In order to address these issues, special attention is paid to the specification stage of the nonlinear model and the investigation of the performance of forecast evaluation measures. The former consists of a battery of recently developed statistical tests and computationally intensive techniques. While, the examination of the small sample properties of several forecast evaluation measures is implemented through extensive Monte Carlo simulations.

In Chapters 2 and 3, the equilibrium real exchange rate is assumed constant. However, a variety of theoretical models, such as that of Balassa (1964) and Samuelson (1964), imply a non-constant equilibrium in the real exchange rate and estimates, including proxies for the equilibrium determinants, appear significant (see e.g. Lothian and Taylor, 2008; Hegwood and Papell, 2002; Paya and Peel, 2006a). In this framework, International Real Business Cycle (IRBC) models imply a relationship between real exchange rates and consumption series (see Backus and Smith, 1993; Kollmann, 1995). However, these models have received little (if any) empirical support. This discouraging finding gave rise to what is known as the "Backus and Smith puzzle" or the "consumption real exchange rate anomaly". Chapter 4 examines the role of nonlinear dynamics in the generation of the puzzle and provides further evidence on this empirical regularity. Specifically, linear cointegration methods and nonlinear models are employed on quarterly data for several country pairs. In addition, Generalised Impulse Response Functions are introduced so as to examine the time profile of the impact of shocks on the deviations from the IRBC equilibrium.

Chapter 5 explores a different approach to the explanation of the behaviour of the real exchange rate motivated by the recent gravity literature (Anderson and van Wincoop, 2004; Jacks et al., 2008). In Jacks et al. (2008) a micro-founded measure is derived, that enables the construction of long-span trade costs indices. Using this measure, it is shown that trade costs have changed substantially over time. The crucial implication of this finding is that if trade barriers change over time then so should the "degree" of nonlinearity in real exchange rate series. To this end, two nonlinear real exchange rate models are extended to accommodate time-varying market frictions. Moreover, the implications of the estimated models are discussed and compared with those of models based on constant trade costs.

The last chapter summarises the key results and discusses the contributions of the thesis.

CHAPTER 2

Specifying Smooth Transition Regression Models in the Presence of Conditional Heteroskedasticity of Unknown Form[1]

There are considerable dangers in overemphasising the role of significance tests in the interpretation of the data

— *Sir David Roxbee Cox (1924 –)*

2.1 Introduction

Over the last decades there has been a steadily increasing interest in the development and application of nonlinear time series models. A widely used family of nonlinear models is the Smooth Transition Autoregression (STAR) of Ozaki

[1]Monte Carlo experiments for the present and the following chapters were conducted on the Lancaster High Performance Cluster. We are grateful to the administrator, Mike Pacey, for his assistance.

(1978), Granger and Teräsvirta (1993), and Teräsvirta (1994). By allowing regime dependent behaviour, STAR models appear to parsimoniously capture the nonlinear dependence (in the mean) of many economic and financial time series (see, e.g., van Dijk et al., 2002).

Due to the fact that there are various STAR formulations researchers typically adopt a modelling cycle, which consists of specification, estimation and evaluation stages (Eitrheim and Teräsvirta, 1996). Testing linearity comprises the first step of the specification procedure. Several linearity tests against smooth transition nonlinearity have been proposed in the literature (e.g., Luukkonen et al., 1988; Teräsvirta, 1994; Escribano and Jordá, 1999; González and Teräsvirta, 2006). The most widely used are the Lagrange Multiplier type test of Teräsvirta (1994) and the test derived by Escribano and Jordá (1999). Despite the fact that there is a vast empirical literature suggesting that the residuals of many regression models in economics and finance exhibit time-varying conditional variance (Engle, 1982, 2001), the robustness of these tests to conditional heteroskedasticity has not been thoroughly addressed.

As noted by a number of researchers neglected heteroskedasticity may result in substantial oversizing of linearity tests. It also holds that the performance of tests for conditional heteroskedasticity depends on the correct specification of the conditional mean (see, e.g., Blake and Kapetanios, 2007, and references therein). Notably, Granger and Teräsvirta (1993) argue that the Autoregressive Conditional Heteroskedastic (ARCH) model of Engle (1982) although linear in mean can complicate tests for linearity. Wong and Li (1997) show through Monte Carlo simulations that tests for Threshold Autoregression (TAR) assuming a constant conditional variance can be heavily oversized in the presence of ARCH innovations. A similar empirical finding is provided by Hurn and Becker (2007) for the neural network test of Teräsvirta et al. (1993). Further, Bera and Higgins (1997) argue that bilinear processes can be confused with ARCH processes due to the similarity of their unconditional moment structure.

Granger and Teräsvirta (1993), based on the work of Davidson and MacKinnon (1985), propose a robust test for linearity against STAR nonlinearity in the presence of unknown form of heteroskedasticity. However, Lundbergh and Teräsvirta (1998) illustrate that although the above robustification significantly reduces oversizing it may result in a severe loss of power. To this end, they suggest using the original test and examining the presence of neglected heteroskedasticity in the following steps of the modelling procedure. However, such a modelling cycle may often lead to the misspecification of the conditional mean.

In this chapter, we investigate the effect of conditional heteroskedasticity on the linearity test of Escribano and Jordá (1999) as well as four heteroskedasticity robust versions. The first three utilise the Heteroskedasticity Consistent Covariance Matrix Estimators (HCCMEs) considered in White (1980) and MacKinnon and White 1985, while the last one employs the Fixed Design Wild Bootstrap of Kreiss (1997) and Gonçalves and Kilian (2004). HCCMEs are typically employed by researchers due to their asymptotic validity in the presence of heteroskedasticity of unknown form, simple implementation and little computational cost (Long and Ervin, 2000) compared to bootstrap methods. However, in finite samples HCCMEs can be severely biased and, in many cases, they are outperformed by bootstrap methods (Flachaire, 2005). Although we focus on the Generalised Autoregressive Conditional Heteroskedastic (GARCH) model of Bollerslev (1986), we also report results for the Asymmetric GARCH model of Engle (1990), the Exponential GARCH model of Nelson (1991), the GJR GARCH model of Glosten et al. (1993) and the stochastic volatility model advocated by Taylor (1986) and Shephard (1996).

Our findings illustrate that conventional tests may seriously overreject the null of linearity when the null is true and the conditional variance of the error term is time-varying. Further, the degree of oversizing is much higher than the one reported by Lundbergh and Teräsvirta (1998) for the Teräsvirta (1994) test and tends to increase (in many cases rapidly) with the sample size. On the other hand, if the true process is nonlinear in the mean, conditional heteroskedasticity can frequently result in choosing misspecified nonlinear models. Consequently, this can pose problems in the estimation stage of STAR models.

In general, robust tests based on HCCMEs perform poorly. These tests do not always lead to an improvement in empirical size and, usually, result in very low size adjusted power. The final inference technique, the Fixed Design Wild Bootstrap, is superior with respect to all the criteria employed in this study. First, the empirical size of the tests is very close to the nominal significance level. Second, the empirical power is much higher than the rest of the methods. Finally, it results in the selection of correctly specified models in the majority of cases.

The rest of the chapter is organised as follows. Section 2.2 outlines the basic STAR representation, which facilitates the analysis of testing linearity against STAR nonlinearity in Section 2.3. Dealing with conditional heteroskedasticity of unknown form using HCCMEs and the Fixed Design Wild Bootstrap is discussed in Section 2.3.1. The next section investigates the finite sample performance of the tests through Monte Carlo simulations. Section 2.5 presents an empirical application on empirical data. Finally, the last section concludes.

2.2 Smooth Transition Regression Models

The basic STAR model representation for a univariate time series $\{y_t\}$ is given by

$$\begin{aligned} y_t &= \pi_{1,0} + \pi_{1,1} y_{t-1} + \cdots + \pi_{1,p} y_{t-p} + (\pi_{2,0} + \\ &\quad + \pi_{2,1} y_{t-1} + \cdots + \pi_{2,p} y_{t-p}) F(s_t; \gamma, c) + \epsilon_t, \qquad t = 1, \ldots, T, \end{aligned} \tag{2.1}$$

or equivalently

$$y_t = \boldsymbol{\pi}_1' \boldsymbol{x}_t + \boldsymbol{\pi}_2' \boldsymbol{x}_t F(s_t; \gamma, c) + \epsilon_t, \qquad t = 1, \ldots, T, \tag{2.2}$$

where $\boldsymbol{x}_t = (1, \tilde{\boldsymbol{x}}_t')'$ with $\tilde{\boldsymbol{x}}_t = (y_{t-1}, \ldots, y_{t-p})'$ and $\boldsymbol{\pi}_j = (\pi_{j,0}, \ldots, \pi_{j,p})'$, for $j = 1, 2$. The STAR model can be easily extended to a Smooth Transition Regression (STR) model by augmenting Equation (2.2) with exogenous regressors. Hence, our analysis can be generalised to the STR model in a straightforward manner. Depending on the derivation of the linearity test under consideration, it is assumed that the error term, ϵ_t, is either an independent, identically normally distributed random variable, $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$, or a martingale difference sequence. That is, $E[\epsilon_t | \mathcal{I}_{t-1}] = 0$, where $\mathcal{I}_{t-1}$ is the information set up to time $t-1$ consisting of all lagged values of y. Note that in the latter case the variance of the error term is not restricted to be constant. Models that capture the dependence both in the conditional mean and the conditional variance can be found in Lundbergh and Teräsvirta (1998) and Chan and McAleer (2002).

The transition function $F(\cdot)$ is at least fourth-order, continuously differentiable with respect to γ and is bounded between 0 and 1. The selection of the transition function specifies the two common forms of the STAR model. For the Exponential STAR (ESTAR) the transition function is given by

$$F(s_t; \gamma, c) = 1 - \exp\left(-\gamma\left(s_t - c\right)^2\right), \qquad \gamma > 0, \tag{2.3}$$

while for the Logistic STAR (LSTAR),

$$F(s_t; \gamma, c) = \left[1 + \exp\left(-\gamma\left(s_t - c\right)\right)\right]^{-1}, \qquad \gamma > 0, \tag{2.4}$$

where c is a constant and s_t is the transition variable. The transition variable is usually set equal to the lagged endogenous variable y_{t-d}, where the delay parameter d is a positive integer. For $s_t = y_{t-d}$ and $c = \pi_{2,0} = 0$ the ESTAR model collapses to the Exponential Autoregressive (EAR) model of Haggan and Ozaki (1981). Other choices are also possible for the transition variable, such as exogenous variables, nonlinear functions of y_{t-d} or time trends (see, e.g., van Dijk et al., 2002; Paya et al., 2003). The ESTAR transition function is symmetric around

$(s_t - c)$ and admits the limits

$$F(\cdot) \rightarrow 1 \quad \text{as} \quad |s_t - c| \rightarrow +\infty, \tag{2.5}$$

$$F(\cdot) \rightarrow 0 \quad \text{as} \quad |s_t - c| \rightarrow 0. \tag{2.6}$$

While the logistic transition function is asymmetric around $(s_t - c)$ and admits the limits

$$F(\cdot) \rightarrow 1 \quad \text{as} \quad (s_t - c) \rightarrow +\infty, \tag{2.7}$$

$$F(\cdot) \rightarrow 0 \quad \text{as} \quad (s_t - c) \rightarrow -\infty. \tag{2.8}$$

The smoothness parameter $\gamma \in (0, \infty)$ determines the speed of transition of $F(\cdot)$ towards the inner or outer regime and, therefore, the "degree" of nonlinearity (see Figure 2.1). As $\gamma \rightarrow 0$ both transition functions approach a constant and the models become linear. For the ESTAR model the same holds when $\gamma \rightarrow \infty$. Therefore, STAR models nest linear AR models. Moreover, the LSTAR model nests the Threshold Autoregressive (TAR) model with two regimes since for $\gamma \rightarrow \infty$ the logistic transition function approaches the indicator function.

The properties of STR and STAR models are very appealing in modelling nonlinear economic and financial time series. For example, the fact that macroeconomic time series as well as their relationships may be characterised by asymmetries associated with the stages of the business cycle (see, e.g., Skalin and Teräsvirta, 1999; Sensier et al., 2002; Deschamps, 2008) makes Logistic STR models particularly applicable. On the other hand, factors such as market frictions, the sunk costs of international arbitrage as well as heterogeneous agents, may induce nonlinear and symmetric adjustment of many macroeconomic and financial series (e.g., real exchange rates, long gilt futures, dividend-price ratios)

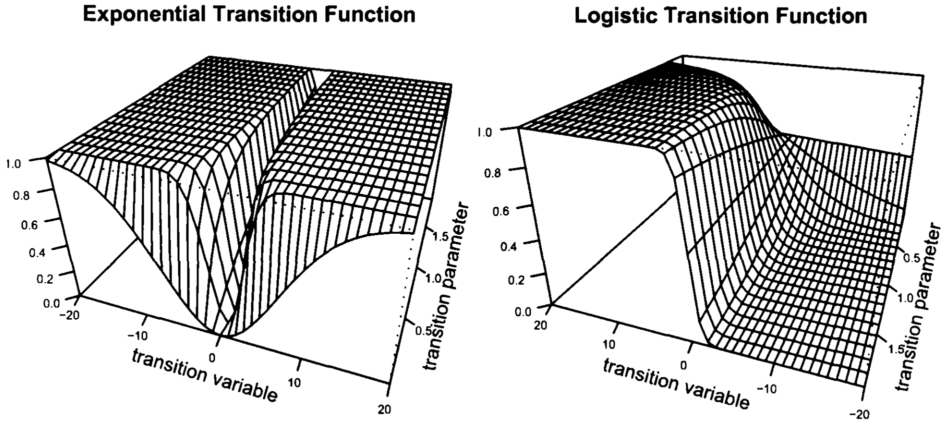

Figure 2.1: The Logistic and Exponential Transition Functions for $\gamma \in \{0.01, \ldots, 2\}$, $s_t \in \{-20, \ldots, 20\}$ and $c = 0$.

motivating the use of Exponential STR models (e.g., Michael et al., 1997; Gallagher and Taylor, 2001; McMillan and Speight, 2002).

2.3 Testing Linearity against Smooth Transition Nonlinearity

There is usually uncertainty about the exact Data Generating Process (DGP) of a variable. Data driven methods allow the selection between competing models and, therefore, provide evidence on the validity of the implications of theoretical models. Several testing procedures have been proposed in the literature to examine whether a series exhibits STAR-type nonlinearity and, in turn, if the nonlinearity displayed is of ESTAR or LSTAR form (e.g., Luukkonen et al., 1988; Teräsvirta, 1994; Escribano and Jordá, 1999; González and Teräsvirta, 2006).

Testing for the nonlinear part of Equation (2.2) gives rise to an nuisance parameter problem (Davies, 1977, 1987). The null hypothesis of linearity corresponds to both H_0: $\pi_2' = 0$ and H_0: $\gamma = 0$. In the former case the parameters γ and c are not identified under the null. While in the latter parameters π_2' and c are not identified. Consequently, classical Lagrange Multiplier (LM) and Wald statistics may not follow standard distributions. Luukkonen et al. (1988) suggest replacing the transition function by a first-order Taylor-series approximation around $\gamma = 0$.[2] This re-parameterisation resolves the identification problem since it does not involve nuisance parameters. The auxiliary regression is given by

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \boldsymbol{x}_t s_t + \boldsymbol{\delta}_2' \boldsymbol{x}_t s_t^2 + u_t, \tag{2.9}$$

where $u_t = \epsilon_t + \mathrm{R}(\gamma, s_t)$, $R(\cdot)$ is the remainder term of the Taylor series. However, if $s_t = y_{t-d}$ and $d \leq p$ then

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \tilde{\boldsymbol{x}}_t s_t + \boldsymbol{\delta}_2' \tilde{\boldsymbol{x}}_t s_t^2 + u_t, \tag{2.10}$$

so as to avoid perfect multicollinearity among the explanatory variables. In order to ease notation we assume $p < d$. The null hypothesis of linearity becomes H_0: $\boldsymbol{\delta}_1' = \boldsymbol{\delta}_2' = 0$. Under the null, the LM test statistic has an an asymptotic χ^2 distribution with the degrees of freedom equal to the number of restrictions. A drawback of the above auxiliary regression arises for LSTAR processes ($\boldsymbol{\delta}_2' = 0$). In particular, if y_t is an LSTAR process and only intercept changes are significant across regimes then the nonlinearity test will lack power (see, e.g., Escribano and Jordá, 2001). To this end, the authors suggest using a third order Taylor series approximation of the logistic function. This yields the auxiliary regression

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \boldsymbol{x}_t s_t + \boldsymbol{\delta}_2' \boldsymbol{x}_t s_t^2 + \boldsymbol{\delta}_3' \boldsymbol{x}_t s_t^3 + u_t. \tag{2.11}$$

Teräsvirta (1994) proposes a modelling procedure based on Equation (2.11)

1. Specification of a linear model. The selection of the lag order can be implemented by using either a criterion such as the Akaike Information Criterion (AIC) or significance tests.

2. Testing the null hypothesis of linearity, H_{00}: $\boldsymbol{\delta}_1' = \boldsymbol{\delta}_2' = \boldsymbol{\delta}_3' = 0$. Often,

[2] Note that test based on Taylor-series approximations do not have direct power against a single alternative.

the transition variable is set equal to the lagged endogenous variable y_{t-d}. However, there may be uncertainty about the appropriate delay parameter, d, in the STR model. In this case, we can determine the transition variable by testing H_{00} for various values of d and selecting the one for which the p-value is smallest.

3. Selecting the transition function. The choice between ESTAR and LSTAR models can be based on the following sequence of null hypotheses:

$$\begin{aligned} H_{03}: \boldsymbol{\delta}_3' &= 0, \\ H_{02}: \boldsymbol{\delta}_2' &= 0 \mid \boldsymbol{\delta}_3' = 0, \\ H_{01}: \boldsymbol{\delta}_1' &= 0 \mid \boldsymbol{\delta}_2' = \boldsymbol{\delta}_3' = 0. \end{aligned}$$

If the p-value for the F-test of H_{02} is smaller than that for H_{01} and H_{03} then we select the ESTAR family, otherwise we choose the LSTAR family.

Whilst, Teräsvirta (1994) uses a third-order Taylor expansion of the logistic transition function and a first-order Taylor expansion for the exponential function, Escribano and Jordá (1999) augment the regression equation with a second-order expansion of the exponential function. Note that even (odd) powers of the Taylor approximation of the logistic (exponential) function are all zero. The point of using a second-order Taylor expansion lies in the fact that the logistic function has one inflection point while the exponential possesses two. The auxiliary regression is given by

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \boldsymbol{x}_t s_t + \boldsymbol{\delta}_2' \boldsymbol{x}_t s_t^2 + \boldsymbol{\delta}_3' \boldsymbol{x}_t s_t^3 + \boldsymbol{\delta}_4' \boldsymbol{x}_t s_t^4 + u_t. \tag{2.12}$$

Escribano and Jordá (1999) claim that this procedure improves the power of both the linearity test and the selection procedure test. The null hypothesis of linearity corresponds to $H_0^1 : \boldsymbol{\delta}_1' = \boldsymbol{\delta}_2' = \boldsymbol{\delta}_3' = \boldsymbol{\delta}_4' = 0$. Under this null the test statistic has asymptotically a χ^2 distribution with $4(p+1)$ degrees of freedom. In finite samples, however, the χ^2 test can be oversized. To this end, the F version is preferred because it has better small size properties. The selection procedure between ESTAR and LSTAR changes to

1. Test the null hypothesis $H_0^L : \delta_2' = \delta_4' = 0$, with an F-test, (F_L).

2. Test the null hypothesis $H_0^E : \delta_1' = \delta_3' = 0$, with an F-test, (F_E).

3. If the p-value of F_L is lower than F_E then select an ESTAR. Otherwise, select an LSTAR.

The use of the F-test is based on the assumption that the error term in Equation (2.2) is independent, identically and normally distributed. However, the assumption of constant conditional variance may be too strict when it comes to empirical applications.

2.3.1 Dealing with Conditional Heteroskedasticity

Since the work of Engle (1982) and Bollerslev (1986) it has become a stylized fact that the residuals of many dynamic regression models exhibit conditional heteroskedasticity. The evidence of conditional heteroskedasticity becomes overwhelming as we move from low frequencies of data (annual, quarterly) to high frequencies (monthly,weekly, daily) and especially ultra high frequencies (five minutes, tick-by-tick) (see, e.g., Dacorogna et al., 2001).

Applications of STAR models and, therefore, of the corresponding linearity tests cover all possible frequencies. Notably, Skalin and Teräsvirta (1999) investigate the properties of the Swedish business cycle by fitting STAR models to annual macroeconomic time series, which cover the period 1861 to 1988. Long spans of annual data are also employed in studies examining the presence of nonlinearities in real exchange rates (Lothian and Taylor, 2008; Paya and Peel, 2006a). Gallagher and Taylor (2001) investigate the risky arbitrage hypothesis by fitting an ESTAR-ARCH model to quarterly data on the U.S. market log dividend-price ratio. Further, Taylor et al. (2001), Kilian and Taylor (2003) and Paya et al. (2003)

show that ESTAR models can capture the behaviour of quarterly and monthly real exchange rates in the post-Bretton Woods era. A similar conclusion is derived for the futures basis of the S&P 500 and the FTSE 100 by Monoyios and Sarno (2002), who use daily data. A model that allows simultaneous modelling of the first and second moments is the STAR-Smooth Transition GARCH (STAR-STGARCH) introduced by Lundbergh and Teräsvirta (1998). The model is applied to two daily series, the Swedish OMX index and the Japanese yen U.S. dollar exchange rate. In a related study, Chan and McAleer (2002) investigate the statistical properties of the STAR-GARCH model and fit the model to the S&P 500 daily returns. Taylor et al. (2000) examine arbitrage opportunities in the FTSE 100 using 1,2 and 5 minutes frequency data. The authors adopt an Exponential Smooth Transition Error Correction model to obtain transactions costs and trade speeds faced by arbitrageurs who exploit mispricing of FTSE 100 futures contracts relative to spot prices. Their results indicate significant ARCH type heteroskedasticity in the estimated residuals.

Linearity tests against smooth transition nonlinearity are implemented in most of the above studies. The question that naturally arises is whether these tests are robust to a time-varying conditional variance and, if not, whether there are ways of robustification.

In this study, we focus on the Escribano and Jordá (1999) test and adopt a nonparametric approach to deal with conditional heteroskedasticity of unknown form in Equation (2.12). The use of parametric models requires knowledge of the type and the precise form of conditional heteroskedasticity. However, it is unlikely that such information is available in practice. Therefore, we examine the performance of the HCCME of White (1980), two HCCMEs examined by MacKinnon and White (1985), and, finally, the Fixed Design Wild Bootstrap of Kreiss (1997) and Gonçalves and Kilian (2004, 2007).

2.3.2 Hypothesis Testing

A general representation for all the linear auxiliary regressions of the previous section is given by

$$y_t = \boldsymbol{\delta}' \boldsymbol{z}_t + u_t. \tag{2.13}$$

For the Escribano and Jordá (1999) test $\boldsymbol{\delta} = (\boldsymbol{\delta}_0', \ldots, \boldsymbol{\delta}_4')'$ and $\boldsymbol{z}_t = (\boldsymbol{\zeta}_{0,t}', \ldots, \boldsymbol{\zeta}_{4,t}')'$ with $\boldsymbol{\zeta}_{j,t} = \boldsymbol{x}_t s_t^j$, for $j = 0, \ldots, 4$. The null hypothesis of linearity, ESTAR or LSTAR can be written as H_0: $\boldsymbol{R}\widehat{\boldsymbol{\delta}} = 0$, where $\boldsymbol{R}$ is the $q \times 5(p+1)$ selector matrix with q denoting the number of restrictions. Testing for linearity requires $4(p+1)$ restrictions while for the ESTAR and LSTAR $2(p+1)$. The Wald form of the test statistic can be written as

$$W = \left(\boldsymbol{R}\widehat{\boldsymbol{\delta}}\right)' \left(\boldsymbol{R}\widehat{\boldsymbol{\Psi}}\boldsymbol{R}'\right)^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\delta}}\right), \tag{2.14}$$

where $\widehat{\boldsymbol{\Psi}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\widehat{\boldsymbol{\Omega}}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ denotes the covariance matrix of the estimates $\widehat{\boldsymbol{\delta}}$. Consistency of the estimator $\widehat{\boldsymbol{\Psi}}$ is required when drawing inferences. Assuming that the residuals, u_t, are independent, identically and normally distributed with variance σ_u^2 yields

$$\text{LS} : \widehat{\boldsymbol{\Omega}} = \widehat{\sigma}_u^2 \boldsymbol{I}, \tag{2.15}$$

where $\boldsymbol{I}$ is the identity matrix. In this case, W/q is F distributed under the null.

However, in the presence of heteroskedasticity the diagonal elements of $\widehat{\boldsymbol{\Omega}}$ will not be constant. It follows that the ordinary least squares estimator of the covariance matrix (LS) will be biased and conventional tests will generally have non-standard distributions (e.g., Flachaire, 2005; Long and Ervin, 2000). That is, the Wald statistic will not follow an F distribution, even asymptotically. In this case, HCCMEs are usually employed by researchers.[3] Eicker (1963) and White

[3]Although we focus on the presence of heteroskedastic errors, serial correlation may also be present in real world applications. In that case, heteroskedasticity and autocorrelation covariance estimators can be employed.

(1980) propose the following heteroskedasticity consistent estimator

$$\text{HC0}: \ \widehat{\Omega} = \text{diag}(\widehat{u}_t^2), \tag{2.16}$$

which allows asymptotic inference. The idea is to use $\widehat{u}_t^2$ to estimate the variance of the error term at time t. Unfortunately, the HC0 and F-tests can be heavily biased in finite samples. To this end, MacKinnon and White (1985), based on the work of Hinkley (1977), Horn et al. (1975) and Efron (1982), consider three alternative HCCMEs. The two estimators employed in this study are

$$\text{HC2}: \ \widehat{\Omega} = \text{diag}\left(\frac{\widehat{u}_t^2}{1-h_{tt}}\right), \tag{2.17}$$

$$\text{HC3}: \ \widehat{\Omega} = \text{diag}\left(\frac{\widehat{u}_t^2}{(1-h_{tt})^2}\right), \tag{2.18}$$

where $h_{tt} = z_t(\boldsymbol{Z}'\boldsymbol{Z})^{-1}z_t'$ is the t^{th} diagonal element of the "hat" matrix. The authors show that both HC2 and HC3 lead to a marked improvement in small samples. Further, Long and Ervin (2000) suggest using HC3 when the sample size is less than 250 observations. Despite the fact that the latter estimators are superior to HC0, they too are biased.

The fact that Wald tests do not follow F distributions, even asymptotically, as well as the poor finite size properties of HCCMEs motivate the use of bootstrap methods for conducting statistical inference. The rationale of bootstrap methods is to approximate the finite sample distribution of the test statistic under the null by simulation. In general, bootstrap tests may lead to a significant improvement in terms of the Error in Rejection Probability (ERP) (see, e.g, Davidson and MacKinnon, 1999). The findings of Beran (1988) indicate that the ERP of a bootstrap test is of lower order, in general $O(T^{-0.5})$, than the asymptotic tests when the test statistic is asymptotically pivotal. Moreover, Davidson and MacKinnon (1999) illustrate that a further refinement of the same order occurs when the test statistic is independent of the bootstrap DGP. It follows that, in many cases, bootstrap tests are more precise than asymptotic tests by $O(T^{-1})$.

A bootstrap technique which deals with heteroskedasticity of unknown form is the Wild Bootstrap. The asymptotic validity of the Wild Bootstrap for linear regressions is established in Wu (1986), Liu (1988) and Mammen (1993). Kreiss (1997) and Gonçalves and Kilian (2004) extend the analysis to stationary autoregressions with conditional heteroskedastic errors. As far as linearity tests are concerned, Hurn and Becker (2007) illustrate that the Wild Bootstrap improves upon the neural network test of Teräsvirta et al. (1993) when there is GARCH type conditional heteroskedasticity in the residuals.

We now describe the Fixed Design Wild Bootstrap procedure for testing the hypothesis of linearity, ESTAR nonlinearity or LSTAR nonlinearity

1. Estimate Equation (2.13) and compute the F-statistic, $\tilde{F}$.

2. Estimate the restricted model and obtain the estimated coefficient vector $\widehat{\boldsymbol{\delta}}_r$ and the restricted residuals $\widehat{u}_{r,t}$.

3. Generate B "fake" series according to null DGP

$$y_t^b = \widehat{\boldsymbol{\delta}}_r' \boldsymbol{z}_t + \epsilon_t^b,$$

where the residuals ϵ_t^b are constructed by multiplying the estimated restricted residuals $\widehat{u}_{r,t}$ by a random variable η_t.The η_t must be mutually independent drawings from a distribution independent of the original data with mean 0 and variance 1. Liu (1988) and Davidson and Flachaire (2001) suggest using the Rademacher distribution

$$\eta_t = \begin{cases} -1 & \text{with probability } p = 0.5\ , \\ +1 & \text{with probability } (1-p). \end{cases}$$

The Rademacher distribution has the properties $E[\eta_t] = 0$, $E[\eta_t^2] = 1$, $E[\eta_t^3] = 0$, and $E[\eta_t^4] = 1$. A consequence of these properties is that any heteroskedasticity or symmetric non normality in the estimated residuals

$\widehat{u}_{r,t}$ is preserved in the newly created residuals. The Wild Bootstrap matches the moments of the observed error distribution around the estimated regression function at each design point, $\widehat{y}_t^b$. Liu (1988) and Mammen (1993) show that the asymptotic distribution of the Wild Bootstrap statistics are the same as the statistics they try to mimic.

4. Regress each "fake" series $\boldsymbol{y}^b$ on $\mathbf{Z}$ and compute the F-statistic, $\tilde{F}_b$, so as to obtain the empirical distribution for the F-statistic under the null.

5. Compute the p-value as the percentage of times the simulated statistic $\tilde{F}_b$ is more extreme than the original statistic $\tilde{F}$

$$p_b = \frac{1}{B}\sum_{b=1}^{B} I(\tilde{F} \leq \tilde{F}_b)$$

   where $I(A)$ is the indicator function, which takes the value of 1 if event A occurs and 0 otherwise.

6. Reject the null if p_b is smaller than the chosen significance level.

In the next section, we conduct Monte Carlo simulation exercises in order to examine the accuracy of the inference procedures under different error processes and sample sizes.

2.4 Monte Carlo Simulation

As aforementioned, the LM test of Teräsvirta (1994) performs poorly, in terms of size, when there is conditional heteroskedasticity. On the other hand, the robust version proposed by Granger and Teräsvirta (1993) appears to lack power (Lundbergh and Teräsvirta, 1998). In this section, we investigate whether there is a similar effect on the Escribano and Jordá (1999) test and the performance of the heteroskedasticity robust inference techniques.

The simulation exercises focus on a simple STAR(1) conditional mean equation examined by Escribano and Jordá (2001)

$$y_t = \pi_{1,1} y_{t-1} + \pi_{2,1} y_{t-1} F(y_{t-d}; \gamma, c) + \epsilon_t, \qquad t = 1, \ldots, T, \tag{2.19}$$

where $\pi_{1,1} = 0.3$ and $\pi_{2,1} = -0.9$ and $c = 0$. For the error term we adopt various conditional heteroskedastic processes. The first type is the standard GARCH(1,1) proposed by Bollerslev (1986) to capture volatility clustering,

$$\epsilon_t = e_t h_t^{1/2}, \quad h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1}, \quad e_t \sim \mathcal{N}(0,1) \tag{2.20}$$

where h_t denotes the conditional variance at time t. We follow Gonçalves and Kilian (2004) and set $(\alpha, \beta) \in \{(0,0), (0.5, 0), (0.3, 0.65), (0.2, 0.79), (0.05, 0.94)\}$ and $\omega = 1 - \alpha - \beta$, which implies an unconditional variance of unity. We also consider ARCH type models which allow asymmetric effects of positive and negative shocks on volatility (see Bollerslev et al., 1993). In particular, we employ the Exponential GARCH (EGARCH) model of Nelson (1991), the Asymmetric GARCH (AGARCH) of Engle (1990) and the GJR GARCH model proposed by Glosten et al. (1993).

EGARCH:

$$\begin{aligned} \epsilon_t &= e_t h_t^{1/2}, \quad \ln(h_t) = -0.23 + 0.9 \ln(h_{t-1}) + 0.25 \left(e_{t-1}^2 - 0.3 e_{t-1}\right), \\ e_t &\sim \mathcal{N}(0,1). \end{aligned} \tag{2.21}$$

AGARCH:

$$\begin{aligned} \epsilon_t &= e_t h_t^{1/2}, \quad h_t = 0.0216 + 0.6896 h_{t-1} + 0.3174 \left(\epsilon_{t-1} - 0.1108\right)^2, \\ e_t &\sim \mathcal{N}(0,1). \end{aligned} \tag{2.22}$$

GJR GARCH:

$$
\begin{aligned}
\epsilon_t &= e_t h_t^{1/2}, \quad h_t = 0.005 + 0.7h_{t-1} + 0.28\left(\epsilon_{t-1}^2 - 0.23\epsilon_{t-1}\right), \\
e_t &\sim \mathcal{N}(0,1).
\end{aligned} \tag{2.23}
$$

The form of the error processes and the parameter values are based on Engle and Ng (1993). The above models are motivated by the so-called "leverage effect" characterising stock returns. This effect was first noted by Black (1976)

> *"a drop in the value of the firm will cause a negative return on its stock, and will usually increase the leverage of the stock ... That rise in the debt-equity ratio will surely mean a rise in the volatility of the stock."*

An alternative explanation is the asymmetric reaction of asset markets to "good" and "bad" news. Finally, we consider a stochastic volatility model proposed by Taylor (1986) and employed by Shephard (1996) to capture the volatility of returns on the Nikkei index and the Japanese yen and Deutsch mark against the pound sterling.

$$
\begin{aligned}
\epsilon_t &= e_t \exp(h_t), \quad h_t = 0.951h_{t-1} + 0.5e_t, \\
(\epsilon_t, e_t) &\sim \mathcal{N}(0, \mathrm{diag}(0.18, 1)).
\end{aligned} \tag{2.24}
$$

We restrict the experiments to sample sizes of 100, 250, 500, and 1000 observations, which cover the majority of data sets used in applied work. Larger sizes, such as the ones available in ultra high frequency studies, are not examined due to the computationally intensive nature of the experiment. However, our results are indicative of the change of the performance of the tests with the sample size. The nominal significance level is set to 5% and the number of simulated series as well as the number of Wild Bootstrap replications per series is 1000.[4] The first 100 observations are discarded to avoid initialisation effects.

[4] In this case, the overall significance level may differ from the 5% due to multi-step testing.

2.4.1 Empirical Size of Linearity Tests

In order to investigate the size properties of the tests, we set the smoothness parameter γ equal to 0. Hence, Equation (2.19) becomes an AR(1) model with conditional homoskedasticity (when $\alpha = \beta = 0$) or conditional heteroskedasticity. Tables 2.1 and 2.2 report results for the null hypotheses of linearity and the percentage of times an ESTAR model is selected rather than an LSTAR. The percentage of LSTAR selections can be computed by subtracting the percentage of ESTAR selections from the empirical size of the tests. Results for the tests based on the least squares covariance matrix estimator, the three heteroskedasticity consistent covariance matrix estimators and the Wild Bootstrap are presented in the columns labelled LS, HC0, HC2 and HC3, and WB, respectively. In addition, Figure 2.2 provides a visual view of the ERP (the difference between the empirical size and the nominal level of a test) for stationary GARCH processes.

Starting with the standard F version of the Escribano and Jordá test (column LS), several interesting conclusions emerge. First, the test may exhibit serious size distortions. The null of linearity can be rejected up to 81% of the times for a nominal significance level of 5% when the error process is AGARCH or GJR GARCH and $T = 1000$. These size distortions are much more severe than the ones reported in Lundbergh and Teräsvirta (1998) for the Teräsvirta (1994) test. It should be noted that the two simulation experiments differ. The authors examine an AR(4) model with a different GJR-GARCH residual process. Therefore, direct comparisons between the two tests cannot be made. For the GARCH models there is a positive relationship between the degree of oversizing and the value of the ARCH parameter (see Figure 2.2). Second, the bias of the empirical size can rapidly increase with the sample size. Hence, application of the test to large data sets, such as the ones available for daily or intra-daily stock returns and exchange rate returns, is most likely to result in false inference. Finally, it appears that the test does not favour either alternative (ESTAR and LSTAR), which is also true for the remaining inference techniques.

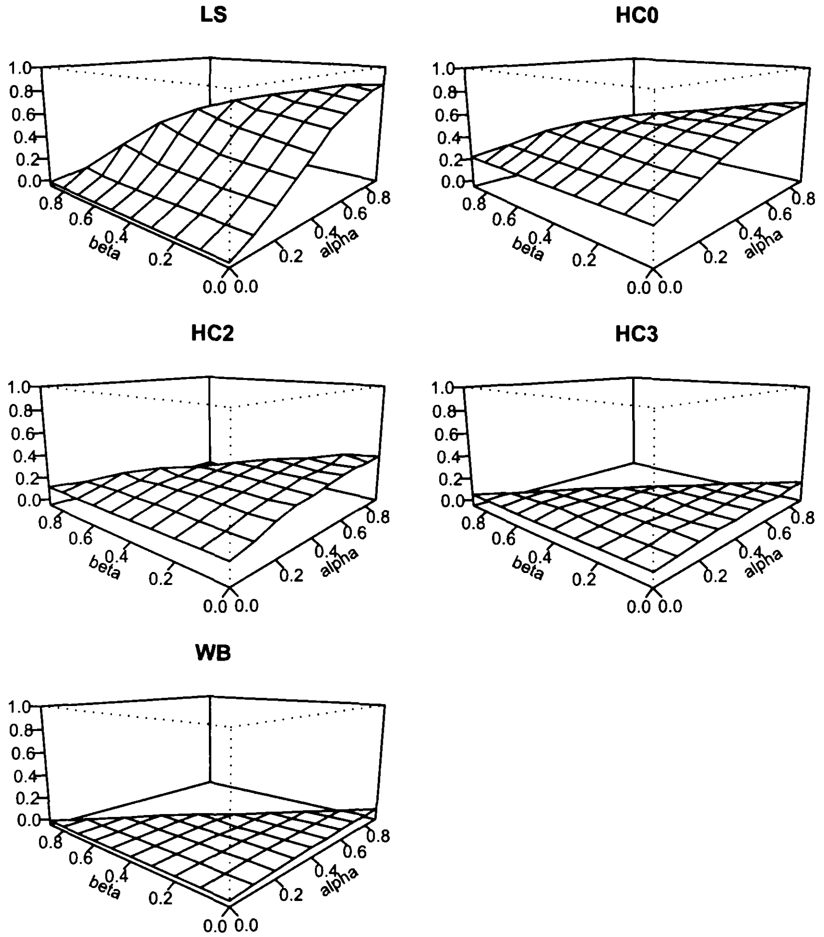

Figure 2.2: Error in rejection probability in LS, HC0, HC2, HC3 and WB linearity tests in the presence of conditional heteroskedasticity. The DGP is an AR(1)-GARCH(1,1) model. The AR coefficient $\phi = 0.3$, and the GARCH parameters $\alpha \in \{0, 0.1, \ldots, 0.8, 0.9\}$ and $\beta \in \{0, 0.1, \ldots, 0.8, 0.9\}$ satisfy $\alpha + \beta < 1$. The unconditional variance of the error process is set to unity ($\omega = 1 - \alpha - \beta$).

Table 2.1: Empirical Size of Wald F-tests

DGP: $y_t = 0.3y_{t-1} + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$,

$h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}$, $e_t \sim \mathcal{N}(0,1)$.

Sample Size $T = 100$

		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	0.05	0.39	0.21	0.09	0.06	0.02	0.19	0.10	0.05	0.03
0.50	0.00	0.37	0.59	0.30	0.13	0.08	0.18	0.29	0.15	0.07	0.04
0.30	0.65	0.28	0.55	0.28	0.12	0.06	0.14	0.27	0.13	0.04	0.03
0.20	0.79	0.20	0.50	0.23	0.09	0.06	0.10	0.23	0.11	0.04	0.03
0.05	0.94	0.07	0.40	0.20	0.10	0.06	0.04	0.19	0.09	0.04	0.03

Sample Size $T = 250$

		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	0.04	0.32	0.18	0.10	0.05	0.01	0.15	0.09	0.04	0.02
0.50	0.00	0.50	0.60	0.32	0.13	0.08	0.25	0.33	0.18	0.06	0.05
0.30	0.65	0.47	0.58	0.30	0.12	0.08	0.22	0.28	0.16	0.05	0.05
0.20	0.79	0.38	0.53	0.29	0.13	0.06	0.18	0.28	0.15	0.06	0.03
0.05	0.94	0.10	0.39	0.21	0.11	0.05	0.05	0.21	0.12	0.05	0.02

Sample Size $T = 500$

		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	0.04	0.28	0.16	0.09	0.06	0.01	0.14	0.08	0.04	0.02
0.50	0.00	0.64	0.63	0.35	0.16	0.08	0.36	0.38	0.22	0.09	0.05
0.30	0.65	0.62	0.60	0.34	0.14	0.07	0.33	0.34	0.20	0.08	0.04
0.20	0.79	0.52	0.51	0.28	0.11	0.05	0.26	0.29	0.16	0.07	0.03

Continued on Next Page...

0.05	0.94	0.11	0.33	0.19	0.10	0.04	0.05	0.20	0.11	0.05	0.02
Sample Size $T = 1000$											
		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	0.06	0.24	0.15	0.11	0.06	0.04	0.14	0.08	0.05	0.03
0.50	0.00	0.70	0.57	0.35	0.13	0.07	0.36	0.36	0.21	0.07	0.05
0.30	0.65	0.72	0.56	0.30	0.14	0.07	0.39	0.35	0.20	0.08	0.04
0.20	0.79	0.66	0.50	0.29	0.14	0.06	0.34	0.28	0.17	0.07	0.03
0.05	0.94	0.18	0.34	0.21	0.12	0.06	0.10	0.21	0.12	0.06	0.03

NOTE: The table reports the empirical size of the LS, HC0, HC2, HC3 and the WB linearity tests, as well as the percentage of times an ESTAR model is selected rather than an LSTAR (ESTAR selection). The nominal significance level is 5%.

Turning to the heteroskedasticity robust tests, we observe a strong resemblance between the properties of HC0 and HC2. Both tests seriously overreject the null hypothesis of linearity even when the errors are homoskedastic. Furthermore, oversizing does not appear to decrease (or increase) as we move to larger sample sizes. It should be noted that HC2 gives substantially better results than HC0. A significant reduction in size distortions is achieved by employing the third HC-CME, HC3. The associated test leads to only moderate oversizing with the empirical size reaching a maximum of 16%. However, tests based on HC3 are outperformed by the Fixed Design Wild Bootstrap. The latter method gives almost always the best results and its empirical size is very close to the nominal level irrespective of the sample size and the error process. In the case of homoskedasticity the performance of the Wild Bootstrap is similar to the LS test.

Table 2.2: Empirical Size of Wald F-tests

AR-EGARCH

DGP: $y_t = 0.3y_{t-1} + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$,

$\ln(h_t) = -0.23 + 0.9\ln(h_{t-1}) + 0.25\left(e_{t-1}^2 - 0.3e_{t-1}\right)$, $e_t \sim \mathcal{N}(0,1)$.

	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.37	0.60	0.30	0.12	0.08	0.18	0.30	0.16	0.06	0.04
250	0.57	0.64	0.34	0.13	0.09	0.27	0.34	0.18	0.06	0.06
500	0.69	0.64	0.35	0.15	0.08	0.35	0.39	0.21	0.08	0.06
1000	0.79	0.63	0.34	0.13	0.07	0.41	0.39	0.20	0.06	0.04

AR-AGARCH

DGP: $y_t = 0.3y_{t-1} + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$,

$h_t = 0.0216 + 0.6896h_{t-1} + 0.3174\left(\epsilon_{t-1} - 0.1108\right)^2$, $e_t \sim \mathcal{N}(0,1)$.

	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.29	0.57	0.26	0.09	0.06	0.14	0.26	0.12	0.04	0.03
250	0.55	0.59	0.30	0.15	0.08	0.26	0.29	0.16	0.08	0.04
500	0.71	0.57	0.31	0.11	0.06	0.35	0.34	0.17	0.06	0.03
1000	0.81	0.57	0.29	0.12	0.05	0.41	0.33	0.16	0.05	0.02

AR-GJR-GARCH

DGP: $y_t = 0.3y_{t-1} + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$,

$h_t = 0.005 + 0.7h_{t-1} + 0.28\left(\epsilon_{t-1}^2 - 0.23\epsilon_{t-1}\right)$, $e_t \sim \mathcal{N}(0,1)$.

	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.29	0.59	0.28	0.11	0.07	0.14	0.27	0.12	0.05	0.04
250	0.53	0.58	0.31	0.13	0.07	0.26	0.32	0.16	0.06	0.04

Continued on Next Page…

Table 2.2: Empirical Size of Wald F-tests(Cont'd.)

500	0.67	0.59	0.35	0.12	0.06	0.33	0.34	0.20	0.06	0.03
1000	0.81	0.57	0.30	0.15	0.07	0.38	0.32	0.17	0.07	0.04
AR-Stochastic-Volatility										
DGP: $y_t = 0.3y_{t-1} + \epsilon_t$, $\epsilon_t = e_t \exp(h_t)$, $h_t = 0.951h_{t-1} + 0.5e_t$, $(\epsilon_t, e_t) \sim \mathcal{N}(0, \text{diag}(0.18, 1))$.										
	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.28	0.59	0.28	0.07	0.06	0.13	0.27	0.13	0.04	0.03
250	0.45	0.59	0.32	0.11	0.07	0.23	0.28	0.16	0.04	0.05
500	0.59	0.59	0.30	0.11	0.06	0.30	0.32	0.15	0.06	0.03
1000	0.71	0.59	0.32	0.13	0.06	0.38	0.34	0.18	0.07	0.03

NOTE: See note to Table 2.1.

2.4.2 Empirical Size Adjusted Power of Linearity Tests

Clearly, LS, HC0, and HC2 based tests are seriously oversized. It follows that their empirical power may take large values, which can, partially, be attributed to the presence of conditional heteroskedasticity. In order to make comparisons between alternative methods meaningful, we adjust for the bias in the empirical size. Empirical size adjusted power is reported for all tests but the Fixed Design Wild Bootstrap, for which no size adjustment is made. This should not have a significant impact on inference, since the empirical size of the Wild Bootstrap is very close to the nominal level. For the power experiments, we set the transition variable equal to y_{t-1} and the transition parameter equal to 1. The rest of the details for the simulation procedure are the same as for the size experiment. Tables

2.3 and 2.4 report the results. We have also examined a LSTAR DGP. The results are qualitatively similar to the ESTAR case and are omitted so as to save space.

Table 2.3: Empirical Size Adjusted Power of Wald F-tests

DGP: $y_t = 0.3y_{t-1} - 0.9y_{t-1}[1 - \exp(-y_{t-1}^2)] + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$,
$h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}$, $e_t \sim \mathcal{N}(0,1)$.

Sample Size $T = 100$

		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	0.23	0.07	0.10	0.14	0.30	0.20	0.05	0.08	0.12	0.26
0.50	0.00	0.02	0.07	0.04	0.09	0.27	0.02	0.05	0.03	0.07	0.24
0.30	0.65	0.10	0.12	0.11	0.13	0.28	0.07	0.07	0.07	0.10	0.25
0.20	0.79	0.17	0.14	0.19	0.17	0.28	0.14	0.10	0.15	0.15	0.25
0.05	0.94	0.26	0.12	0.12	0.17	0.32	0.23	0.08	0.09	0.15	0.29

Sample Size $T = 250$

		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	0.65	0.11	0.22	0.37	0.68	0.64	0.10	0.21	0.36	0.66
0.50	0.00	0.04	0.08	0.08	0.10	0.40	0.03	0.06	0.07	0.09	0.39
0.30	0.65	0.10	0.09	0.10	0.17	0.48	0.08	0.06	0.08	0.16	0.47
0.20	0.79	0.29	0.13	0.22	0.29	0.53	0.27	0.11	0.20	0.29	0.51
0.05	0.94	0.55	0.15	0.21	0.27	0.63	0.53	0.11	0.19	0.26	0.61

Sample Size $T = 500$

		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	0.92	0.42	0.63	0.76	0.92	0.91	0.41	0.62	0.76	0.91
0.50	0.00	0.05	0.07	0.06	0.20	0.57	0.04	0.05	0.05	0.18	0.56
0.30	0.65	0.12	0.06	0.12	0.23	0.56	0.10	0.04	0.10	0.21	0.54

Continued on Next Page...

0.20	0.79	0.45	0.16	0.28	0.43	0.67	0.43	0.14	0.25	0.41	0.66
0.05	0.94	0.79	0.29	0.45	0.62	0.81	0.78	0.27	0.44	0.62	0.80

Sample Size $T = 1000$

		H_0: Linearity					ESTAR selection				
α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
0.00	0.00	1.00	0.93	0.97	0.99	1.00	0.99	0.92	0.97	0.99	0.99
0.50	0.00	0.10	0.06	0.09	0.33	0.70	0.09	0.04	0.08	0.32	0.70
0.30	0.65	0.15	0.10	0.21	0.33	0.63	0.13	0.06	0.19	0.32	0.61
0.20	0.79	0.54	0.17	0.34	0.49	0.73	0.51	0.15	0.32	0.48	0.72
0.05	0.94	0.94	0.40	0.65	0.78	0.94	0.93	0.38	0.64	0.77	0.93

NOTE: The table reports the empirical size adjusted power of the LS, HC0, HC2, HC3 and the WB linearity tests described in Section 2.3, as well as the percentage of times an ESTAR model is selected rather than an LSTAR (ESTAR selection). The nominal significance level is 5%.

Table 2.4: Empirical Size Adjusted Power of Wald F-tests

ESTAR-EGARCH

DGP: $y_t = 0.3y_{t-1} - 0.9y_{t-1}[1 - \exp(-y_{t-1}^2)] + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$,

$\ln(h_t) = -0.23 + 0.9\ln(h_{t-1}) + 0.25\left(e_{t-1}^2 - 0.3e_{t-1}\right)$, $e_t \sim \mathcal{N}(0,1)$.

	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.04	0.08	0.05	0.07	0.20	0.03	0.05	0.04	0.05	0.15
250	0.03	0.07	0.06	0.07	0.22	0.02	0.04	0.04	0.06	0.19
500	0.04	0.05	0.07	0.07	0.24	0.02	0.03	0.04	0.05	0.22
1000	0.04	0.05	0.05	0.07	0.26	0.02	0.03	0.03	0.06	0.23

ESTAR-AGARCH

DGP: $y_t = 0.3y_{t-1} - 0.9y_{t-1}[1 - \exp(-y_{t-1}^2)] + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$

Continued on Next Page...

Table 2.4: Empirical Size Adjusted Power of Wald F-tests(Cont'd.)

$h_t = 0.0216 + 0.6896 h_{t-1} + 0.3174 \left(\epsilon_{t-1} - 0.1108\right)^2$, $e_t \sim \mathcal{N}(0,1)$.

	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.08	0.11	0.07	0.17	0.23	0.06	0.06	0.05	0.14	0.20
250	0.07	0.08	0.10	0.15	0.33	0.04	0.05	0.08	0.14	0.32
500	0.08	0.08	0.10	0.17	0.34	0.05	0.04	0.08	0.14	0.31
1000	0.07	0.08	0.09	0.19	0.33	0.04	0.04	0.06	0.17	0.30

ESTAR-GJR-GARCH

DGP: $y_t = 0.3 y_{t-1} - 0.9 y_{t-1}[1 - \exp(-y_{t-1}^2)] + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$,

$h_t = 0.005 + 0.7 h_{t-1} + 0.28 \left(\epsilon_{t-1}^2 - 0.23 \epsilon_{t-1}\right)$, $e_t \sim \mathcal{N}(0,1)$.

	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.12	0.07	0.11	0.10	0.20	0.09	0.04	0.08	0.08	0.17
250	0.18	0.10	0.16	0.23	0.42	0.16	0.08	0.15	0.21	0.40
500	0.25	0.16	0.25	0.34	0.62	0.23	0.14	0.23	0.33	0.61
1000	0.38	0.17	0.29	0.45	0.76	0.36	0.14	0.28	0.44	0.75

ESTAR-Stochastic-Volatility

DGP: $y_t = 0.3 y_{t-1} - 0.9 y_{t-1}[1 - \exp(-y_{t-1}^2)] + \epsilon_t$, $\epsilon_t = e_t \exp(h_t)$,

$h_t = 0.951 h_{t-1} + 0.5 e_t$, $(\epsilon_t, e_t) \sim \mathcal{N}(0, \mathrm{diag}(0.18, 1))$.

	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.09	0.09	0.07	0.13	0.19	0.06	0.04	0.05	0.10	0.15
250	0.06	0.06	0.05	0.08	0.20	0.03	0.03	0.03	0.06	0.17
500	0.07	0.06	0.06	0.07	0.16	0.04	0.04	0.04	0.05	0.13
1000	0.05	0.07	0.06	0.10	0.17	0.03	0.03	0.03	0.06	0.14

NOTE: See note to Table 2.3.

A broad tendency that emerges is that the performance of all tests depends crucially on the type of conditional heteroskedasticity. Tests based on the HCCMEs and the ordinary least squares covariance matrix, generally, perform poorly in the presence of conditional heteroskedasticity, with none of them being superior to the others. Furthermore, in many cases these methods have virtually no power to discriminate between linear and nonlinear in mean processes.

The Fixed Design Wild Bootstrap is by far the best method. Its superiority becomes evident in the presence of time-varying conditional variance. For the majority of conditional heteroskedastic processes its power is relatively high and increases with the sample size. While in the case of homoskedasticity its performance is similar or better than the F-test. Unfortunately, the ability of the Wild Bootstrap to detect nonlinearity in the mean is not always satisfactory. For the stochastic volatility process the power of the Fixed Design Wild Bootstrap is extremely low (less than 20%), irrespective of the sample size. Hence, there are cases where all inference techniques perform poorly.

2.4.3 Nonlinear Model Specification

So far we have assumed that the transition variable, or equivalently the delay parameter, is known. However, in real world application the transition variable has to be determined from the data. The selection of a misspecified model is very likely to pose problems in the subsequent stage of estimation. Teräsvirta (1994), inspired by the work of Tsay (1989) on TAR models, suggests choosing the delay parameter that minimises the p-value of the linearity test. The basic idea behind this approach is that on average the power of a correctly specified model should be higher than the power of a misspecified one.

In the last simulation experiments we follow Teräsvirta (1994) and investigate the ability of the tests to identify the correct transition variable. The model design is the same as before, except that we consider three delay parameters, $d = 1, 2, 3$. The same delay parameters specify the candidate transition variables in the linearity tests. We restrict our attention to the GARCH(1,1) process with $\alpha = 0.3$ and $\beta = 0.65$. This choice is motivated by the severe oversizing of the LS and HC tests. Table 2.5 shows the selection frequencies of the transition variables. Note that these are based on the fraction of cases where linearity is rejected. Hence, the results show the probability of choosing the correct delay parameter given linearity is rejected.

Table 2.5: Selection Frequencies of the Delay Parameter, d

DGP: $y_t = 0.3y_{t-1} - 0.9y_{t-1}[1 - \exp(-y_{t-d}^2)] + u_t, u_t = z_t h_t^{1/2}$, $h_t = 0.05 + 0.3u_{t-1}^2 + 0.65h_{t-1}$, $z_t \sim \mathcal{N}(0,1)$.

True Delay Parameter: $d = 1$

T	delay	LS	HC0	HC2	HC3	WB
100	$\mathbf{d = 1}$	0.50	0.21	0.31	0.46	0.72
	$d = 2$	0.30	0.43	0.35	0.30	0.16
	$d = 3$	0.20	0.36	0.34	0.24	0.12
250	$\mathbf{d = 1}$	0.57	0.28	0.39	0.56	0.83
	$d = 2$	0.24	0.39	0.34	0.26	0.09
	$d = 3$	0.19	0.32	0.28	0.18	0.07
500	$\mathbf{d = 1}$	0.57	0.31	0.42	0.59	0.82
	$d = 2$	0.24	0.39	0.28	0.22	0.10
	$d = 3$	0.20	0.30	0.30	0.19	0.08
1000	$\mathbf{d = 1}$	0.60	0.36	0.43	0.57	0.80
	$d = 2$	0.25	0.36	0.31	0.23	0.11
	$d = 3$	0.15	0.28	0.26	0.21	0.08

True Delay Parameter: $d = 2$

T	delay	LS	HC0	HC2	HC3	WB
100	$d = 1$	0.16	0.12	0.15	0.18	0.19
	$\mathbf{d = 2}$	0.67	0.54	0.52	0.56	0.70

Continued on Next Page...

	$d = 3$	0.17	0.33	0.34	0.26	0.10
250	$d = 1$	0.11	0.17	0.17	0.18	0.14
	$\mathbf{d = 2}$	0.75	0.56	0.59	0.65	0.78
	$d = 3$	0.14	0.27	0.24	0.17	0.08
500	d=1	0.15	0.26	0.24	0.23	0.22
	$\mathbf{d = 2}$	0.76	0.50	0.57	0.63	0.74
	$d = 3$	0.09	0.24	0.20	0.14	0.04
1000	$d = 1$	0.27	0.29	0.26	0.25	0.29
	$\mathbf{d = 2}$	0.70	0.51	0.56	0.63	0.69
	$d = 3$	0.04	0.20	0.18	0.12	0.02

True Delay Parameter: $d = 3$

T	delay	LS	HC0	HC2	HC3	WB
100	$d = 1$	0.18	0.12	0.18	0.19	0.17
	$d = 2$	0.23	0.41	0.33	0.25	0.12
	$\mathbf{d = 3}$	0.59	0.47	0.49	0.56	0.71
250	$d = 1$	0.12	0.17	0.15	0.16	0.12
	$d = 2$	0.15	0.38	0.27	0.20	0.06
	$\mathbf{d = 3}$	0.73	0.45	0.58	0.64	0.82
500	$d = 1$	0.12	0.23	0.20	0.20	0.14
	$d = 2$	0.11	0.31	0.22	0.15	0.06
	$\mathbf{d = 3}$	0.77	0.47	0.58	0.65	0.80
1000	$d = 1$	0.19	0.23	0.19	0.19	0.15
	$d = 2$	0.12	0.32	0.24	0.15	0.07
	$\mathbf{d = 3}$	0.69	0.45	0.57	0.67	0.78

NOTE: The table reports selection frequencies of the transition variable y_{t-d}, with $d \in \{1,2,3\}$, when the error term exhibits conditional heteroskedasticity. The chosen delay parameter corresponds to the minimum p-value of the linearity test. True delay parameters are in bold.

Obviously, the use of HC0 and HC2 leads frequently to the selection of misspecified models with HC2 giving again better results. The probability of choosing the wrong transition variable is substantially lower than half when the value of the true delay parameter is one and slightly exceeds half for values two and three. On the contrary, HC3, LS and the Fixed Design Wild Bootstrap appear to perform reasonably well. Overall, the HC3 is outperformed by the ordinary least squares covariance matrix, which is in turn outperformed by the Fixed Design Wild Bootstrap. The difference between the first two methods and the Wild Bootstrap is particularly apparent when the true $d = 1$. The correct selection frequencies for the LS and the HC3 tests vary between 46% and 60%, which implies a high probability of choosing a misspecified model. Whilst for the Wild Bootstrap the corresponding bounds are 73% and 83%. The behaviour of the Wild Bootstrapping is stable across sample sizes and model specifications.

Clearly, the Wild Bootstrap is a valuable technique for testing linearity and, subsequently, specifying STAR models irrespective of the conditional heteroskedasticity of the error process. In the majority of cases it results in valid inferences for the mean equation of a series. To this end, it allows modelling STAR processes when the errors are homoskedastic as well as models which STAR nonlinearity in the mean and conditional heteroskedasticity in the disturbances, such as the STAR-GARCH and the STAR-STGARCH models of Chan and McAleer (2002) and Lundbergh and Teräsvirta (1998), respectively.

2.5 Empirical Applications

The simulation experiments illustrate the likelihood of finding spurious nonlinearity in the mean of economic and financial series when commonly used F-tests are employed and volatility changes occur across time. Since this problem becomes apparent for large sample sizes it would be interesting to apply the linearity tests to empirical data sampled at relatively high frequencies. Therefore, we employ financial time series for which volatility clustering is a well-known fact and high frequency data are available. The presence of time-varying volatility in financial markets has been documented in numerous studies, going back to Mandelbrot (1963) and Fama (1965). Notably, Mandelbrot wrote for stock market returns

> "*... large changes tend to be followed by large changes -of either sign- and small changes tend to be followed by small changes...*".
>
> Mandelbrot (1963, p. 418)

A similar phenomenon is observed for other asset returns, such as exchange rates (Baillie and Bollerslev, 1991, 2002).

However, time-varying volatility is not constrained to high frequency data. The findings of several empirical studies suggest that the volatility of the real exchange rate tends to vary across nominal exchange rate regimes (see, e.g., Mussa, 1986). As a consequence empirical models employing long spans of data typically assume a non constant conditional variance of the error term (see, e.g., Engel and Kim, 1999; Lothian and Taylor, 2008; Paya and Peel, 2006a). To this end, we employ the Lothian and Taylor (1996) two century data set for the dollar-sterling real exchange rate.

A number of theoretical and empirical studies suggest that exchange rate target zones and exchange rate policies, such as "leaning against the wind", may lead to threshold type nonlinearity in the mean of the exchange rate (see, e.g., Krugman, 1991; Lundbergh and Teräsvirta, 2006; Hsieh, 1992). Similarly, factors such as agent heterogeneity, transactions costs or the sunk costs of international arbitrage can induce smooth transition nonlinearity in the the deviation process of asset prices from their fundamental value (Dumas, 1992; Berka, 2002; Kilian and Taylor, 2001). Michael et al. (1997), Taylor et al. (2001) and Kilian and Taylor (2003) among others show that ESTAR models can parsimoniously fit a number of real exchange rates. In the context of stock index futures markets, the findings of Yadav et al. (1994), Dwyer et al. (1996) and Monoyios and Sarno (2002) suggest that TAR and STAR models are capable of explaining the behaviour of the futures basis of major stock indices.

The data set consists of daily closing prices of two stock market indices, namely the Dow Jones and the S&P 500, two nominal exchange rates, the yen-dollar and dollar-sterling, and daily spot and futures prices of the FTSE 100. All series but the last two cover the period from January 1st, 1991 to the December 31st, 2002, which gives a total of 3,131 observations. The data for the spot and future prices of the FTSE 100 span the period January 1st, 1988 to December 31st, 1998, resulting in 2,780 observations. The data were obtained from Datastream. We calculate returns on the Dow Jones, the S&P 500, the dollar-sterling and yen-dollar nominal exchange rates as logarithmic differences of daily closing prices scaled by a factor of 100. Further, we compute the logarithmic FTSE 100 basis b_t according to

$$b_{t,k} = 100 \ln \left(\frac{F_{t,k}}{P_t} \right), \tag{2.25}$$

where $F_{t,k}$ denotes the future price for delivery of the stock at time $k \geq t$ and P_t is the the spot price at time t. Finally, we extend the dollar-sterling real exchange rate (RER) data set of Lothian and Taylor (1996) by using annual data for the U.S. and U.K. consumer price indices and the dollar-sterling nominal exchange rate obtained from the International Financial Statistics database. The extended data set covers the period from 1791 to 2005.

As a preliminary exercise we examine if the series exhibit conditional heteroskedasticity by employing the ARCH LM test derived by Engle (1982). The test is based on the regression equation

$$\widehat{\epsilon}_t^2 = \mu + \sum_{i=1}^{q} a_i \widehat{\epsilon}_{t-i}^2 + v_t, \tag{2.26}$$

where $\widehat{\epsilon}_t$ are the estimated residuals of AR models fitted to the series and μ and a_i, $i = 1, \ldots, q$, are the regression parameters. The lag length of the AR models is determined by using the AIC information criterion for all series but the FTSE 100 basis. For the latter series, we follow Monoyios and Sarno (2002) and set

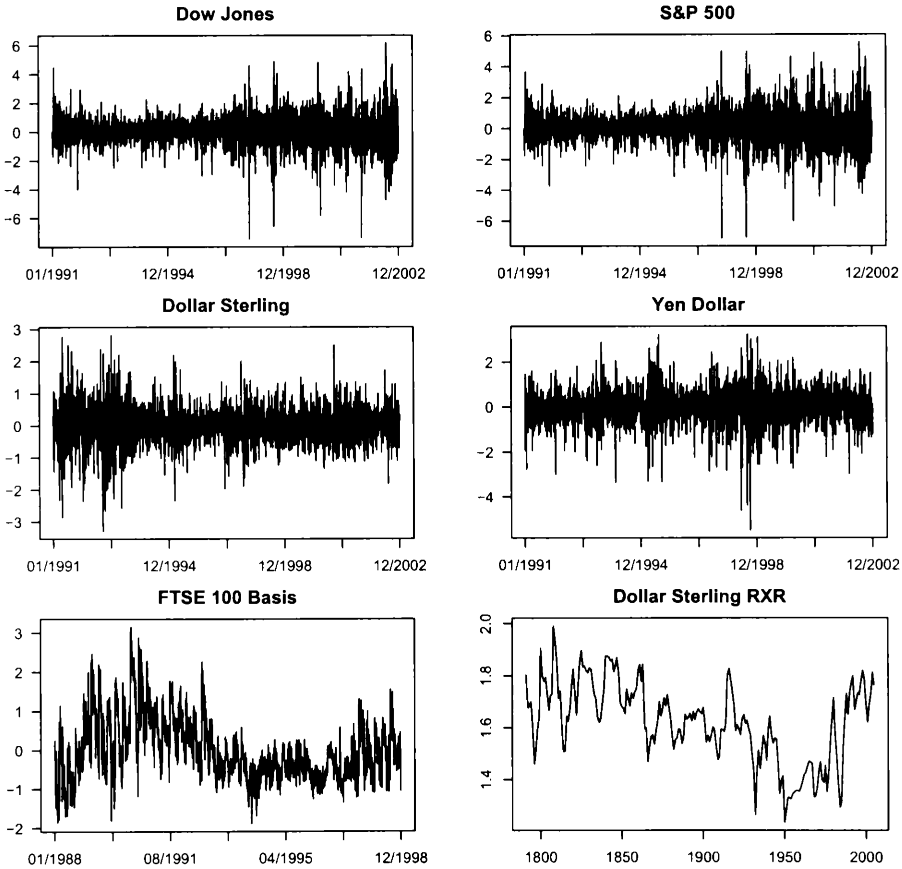

Figure 2.3: Time series plots of empirical data. Daily returns on the Dow Jones and the S&P 500 indices, and the yen-dollar and dollar-sterling nominal exchange rates cover the period January 2nd, 1991 to December 31st, 2002. The basis of the FTSE 100 spans the period January 2nd, 1988 to December 31st, 1998, and the dollar-sterling real exchange rate (RER) the period 1791 to 2005.

the lag length to five. This choice is supported by visual inspection of the partial autocorrelation function. The null hypothesis of no ARCH effects is H_0: $a_i = 0 \; \forall \; i$. Let T denote the sample size, the test statistic given by $T \times R^2$ is asymptotically distributed as χ^2 with q degrees of freedom.

Table 2.6: Results for ARCH LM Tests

Series	χ_1^2	p-value	χ_4^2	p-value
DOW JONES	82.41	0.00	231.61	0.00
S&P 500	134.90	0.00	26.00	0.00
USD STERLING	53.45	0.00	136.00	0.00
YEN USD	39.29	0.00	44.42	0.00
FTSE 100 Basis	28.58	0.00	122.26	0.00
RER	0.03	0.86	0.95	0.92

NOTE: The table reports the χ^2 statistics and the corresponding p-values for ARCH type heteroskedasticity up to orders 1 and 4.

Not surprisingly, Table 2.6 shows that the null hypothesis of no ARCH effects can be rejected at all conventional levels of significance for the high frequency series. Note that at these stage, the rejection of the null hypothesis may be attributed to the presence of STAR type nonlinearity, conditional heteroskedasticity or both. This is due to the fact that like nonlinear in mean tests tend to reject the null in the presence of ARCH effects, ARCH tests also tend to reject the null due to nonlinearities in mean (Blake and Kapetanios, 2007).

Next, we apply the linearity test of Escribano and Jordá (1999) as well as the four robust versions. The choice of the lag order is the same with the one used for the ARCH LM test and the delay parameter is $d = 1, \ldots, 4$. Table 2.7 reports the p-values for the null of linearity corresponding to each transition variable and the selected model.

Table 2.7: Application of Linearity Tests on Empirical Data

Series	Test	H_0: Linearity				Model
		$d=1$	$d=2$	$d=3$	$d=4$	
DOW JONES	LS	**0.000**	0.000	0.000	0.006	LSTAR
	HC0	**0.000**	0.000	0.000	0.000	ESTAR
	HC2	**0.000**	0.063	0.000	0.478	ESTAR
	HC3	0.321	0.988	**0.000**	0.995	LSTAR
	WB	0.317	0.800	0.720	0.990	LINEAR
S&P 500	LS	**0.000**	0.000	0.000	0.006	ESTAR
	HC0	**0.000**	0.000	0.000	0.000	ESTAR
	HC2	**0.000**	0.053	0.000	0.089	ESTAR
	HC3	**0.018**	0.693	0.332	0.987	ESTAR
	WB	0.756	0.539	0.237	0.968	LINEAR
USD	LS	**0.000**	0.000	0.000	0.004	LSTAR
STERLING	HC0	**0.000**	0.000	0.000	0.000	ESTAR
	HC2	**0.000**	0.000	0.000	0.000	LSTAR
	HC3	**0.000**	0.330	0.000	0.451	LSTAR
	WB	0.086	0.487	**0.013**	0.728	LSTAR
YEN USD	LS	**0.000**	0.000	0.000	0.004	LSTAR
	HC0	**0.000**	0.000	0.000	0.000	LSTAR
	HC2	**0.000**	0.000	0.000	0.000	LSTAR
	HC3	0.993	0.585	**0.006**	0.783	LSTAR
	WB	0.961	0.628	0.136	0.710	LINEAR
FTSE 100	LS	**0.000**	0.010	0.071	0.524	LSTAR
BASIS	HC0	**0.000**	0.003	0.000	0.156	LSTAR
	HC2	**0.000**	0.022	0.030	0.314	LSTAR

Continued on Next Page...

Table 2.7: Application of Linearity Tests on Empirical Data(Cont'd.)

	HC3	**0.001**	0.127	0.322	0.548	LSTAR
	WB	**0.000**	0.410	0.491	0.656	LSTAR
RER	LS	0.132	0.782	0.326	0.904	LINEAR
	HC0	0.319	0.870	**0.028**	0.997	ESTAR
	HC2	0.083	0.303	0.749	**0.074**	LSTAR
	HC3	0.228	1.000	0.239	0.904	LINEAR
	WB	**0.043**	0.616	0.462	0.948	ESTAR

NOTE: The table reports p-values of the LS, HC0, HC2, HC3 and WB linearity tests (H_0: Linearity) and the type of STAR nonlinearity selected. Figures in bold denote the selected delay parameter. The nominal significance level is 10%.

Overall, the results are in line with the findings of the simulation experiments. Starting with the returns on the the Dow Jones, the S&P 500 the dollar-sterling and the yen-dollar exchange rate, the Escribano and Jordá (1999) test as well as the HC0 robustification reject the null of linearity for all transition variables. The corresponding marginal significance level is less than 1% in all cases, indicating that the series are characterised by STAR nonlinearity. However, the use of the HC2, HC3 and WB tests results in a substantial decrease in the number of rejections. At the 5% significance level linearity cannot be rejected in 25%, 62.5% and 93.75% of the cases, respectively. Further, there is a wide disparity between the magnitudes of the tests' p-values. An illustrative example is the returns on the yen-dollar exchange rate. For $d = 1$ the p-values of the LS, HC0 and HC2 tests are virtually zero, while the corresponding p-values of the HC3 and WB tests are close to one. The only series for which all methods produce qualitatively similar results with respect to the linearity test is the returns on the dollar-sterling nominal exchange rate.

Turning to the basis of the FTSE 100 there is strong evidence of nonlinearity in mean. At the 5% significance level, the Escribano and Jordá (1999) test indicates nonlinearity for $d = 1, 2$, the HC0 and HC2 based tests for $d = 1, 2, 3$ and the last two tests only for $d = 1$. Overall, the results support setting $d = 1$ since linearity is rejected at all conventional levels of significance irrespective of the test employed. These findings are in line with the theoretical and empirical analysis of Monoyios and Sarno (2002).

As far as the real exchange rate (RER) series is concerned, the HC0 and the Wild Bootstrap tests can reject the null hypothesis of linearity at the 5% significance level. Both tests support the exponential transition function and, hence, symmetric adjustment of the real exchange rate series. For the LS and HC2 tests the smallest p-values are close to the 10% significance, while for HC3 it is substantially larger. Given the results of the ARCH LM test for the dollar-sterling real exchange rate and the superior performance of the Wild Bootstrap, even in the case of homoskedasticity, these findings may be due to the low power of tests based on the HCCMEs when applied to relatively small samples. In addition, nonlinearity tests generally tend not to reject linearity when applied to temporally aggregated nonlinear processes (see, e.g., Granger and Lee, 1999; Paya and Peel, 2006b). Therefore, our findings provide evidence of nonlinearity in the mean of the real exchange rate data.

Overall, the above empirical applications together with the Monte Carlo experiments illustrate the discrepancy between the conclusions drawn using different inference techniques.

2.6 Conclusion

The specification stage of STR models consists of a sequence of tests, which are typically based on the assumption of independent and identically distributed errors. In this chapter, we relaxed this assumption and examined the impact of conditional heteroskedasticity on the tests' performance. We also considered four heteroskedasticity robust versions based on HCCMEs and the Fixed Design Wild Bootstrap.

The findings of the chapter illustrate the dangers of using conventional tests and tests based on HCCMEs. In particular, these tests can exhibit severe size distortions, which increase with the sample size and/or have very low size adjusted power. Further, they frequently lead to the selection of misspecified nonlinear models. Among these methods a HCCME considered by MacKinnon and White (1985) appears to have the best performance. On the other hand, the Fixed Design Wild Bootstrap remedies, at least to a large extend, the deficiencies outlined, allowing inference for both conditional heteroskedastic and homoskedastic errors. Consequently, the application of the Wild Bootstrap provides a valuable alternative to conventional tests.

CHAPTER 3

Forecasting the Behaviour of the Real Exchange Rate using a Long Span of Data

Predictions are hard to make, especially about the future.

Niels Henrik David Bohr (1885 – 1962)

3.1 Introduction

The inception of floating exchange rates in mid-March 1973 was followed by a boom in the interest in explaining the movements of real exchange rates. The observed near unit-root behaviour of the series, however, casted doubts on the predictive ability of the models typically employed until the early 1990s (see, e.g., Taylor and Taylor, 2004, and the references therein). Subsequently, a vast literature, motivated by the presence of frictions in commodity markets, has emerged supporting the existence of a nonlinear adjustment mechanism of the real exchange rate. In accordance with the implications of theoretical models, the findings of numerous empirical studies illustrate that nonlinear models, such as the

Smooth Transition Autoregressive (STAR), provide parsimonious fits to a number of real exchange rates over different time frequencies (e.g., Taylor et al., 2001; Pavlidis et al., 2009a).

Despite the overwhelming evidence supporting the presence of nonlinearites in real exchange rates, the empirical literature on the out-of-sample performance of STAR models is scarce and the question of whether nonlinear models outperform their linear counterparts and the random walk benchmark remains open. One of the few studies on nonlinear real exchange rate forecasting is that of Sarantis (1999). By employing monthly real effective exchange rates for the G-10 countries from 1980 to 1996, the author provides evidence in favour of the presence of significant smooth-transition nonlinear dynamics for the majority of the processes. Moreover, the estimated STAR models provide more accurate forecasts, in terms of the Root Mean Square Error (RMSE) criterion, against the Random Walk (RW) and the Markov Switching model but not the linear autoregressive (AR) model.

A recent study that utilises more sophisticated forecast evaluation techniques and a longer data set for the post-Bretton Wood era is provided by Rapach and Wohar (2006). The authors replicate the results of the seminal papers of Obstfeld and Taylor (1997) and Taylor et al. (2001) by fitting Threshold Autoregressive (TAR) and Exponential STAR (ESTAR) models to four monthly U.S. dollar real exchange rates. Subsequently, they adopt a fixed estimation scheme in order to generate predictions for the following eight years of data. On the basis of point, interval and density forecasts comparisons Rapach and Wohar conclude:

> *"any nonlinearities in monthly real exchange rate data from the post-Bretton Woods period are quite "subtle" for Band-TAR and exponential smooth autoregressive model specifications".*
>
> Rapach and Wohar (2006, p. 341)

These discouraging findings may but do not necessarily imply that the nonlinearity documented in the literature is a spurious artifact. Inoue and Kilian

(2005) illustrate that for linear models in-sample tests tend to have, and in many cases substantially, higher power than out-of-sample tests, which contradicts the conventional view that forecasting comprises the ultimate test of an econometric model. Rossi (2005) also raises concerns regarding the power of out-of-sample predictability tests. Clark and McCracken (2001, 2005a,b) show that commonly used t-type tests, such as that proposed by Diebold and Mariano (1995) and Harvey et al. (1998), may exhibit low power. To this end, Clark and McCracken (2005a) build upon the work of Clark and McCracken (2001) and McCracken (2004) and derive the asymptotic distribution of two F-type tests for the comparison of multi-step forecasts from nested linear models. The tests account for parameter uncertainty and exhibit better power properties than their t-type counterparts. Although their application in this context is appealing, it is not straightforward due to the fact that their derivation is based on the assumption that the regression models are linear in parameters and the processes are stationary. To this end, we relax these assumptions and examine the finite properties of the tests in Section 3.5.

Regarding the comparison of nonlinear with linear AR models, numerous studies suggest that in many cases the in-sample superiority of the former is not accompanied by better predictive ability (see, e.g., Lundbergh and Teräsvirta, 2002; Stock and Watson, 1999). In this framework, power issues turn out to be even more serious.[1]

> *"Many papers exist in which a few series are modeled using a single nonlinear form or class, and usually a good fit is obtained, but often with very little or no improvement in forecasting ability."*
>
> Sir Clive William John Granger (1934–2009)

A possible explanation is that nonlinear models perform better only in specific

[1]The related literature has focused mainly on the comparison of Self-Exciting TAR (SETAR) and AR models. The results presented in Section 3.5 illustrate that this is also the case for STAR models.

states (regime dependent) so that there are windows of opportunity for substantial reduction in prediction errors (Clements, 2005; Boero and Marrocu, 2004). If these occasions are relatively infrequent, then AR models would provide robust forecasts even if the series under consideration is nonlinear (for simulation evidence regarding SETAR models see Clements and Smith, 1999). Put it differently, it would be difficult to identify the gains of forecasting macroeconomic series with nonlinear models, which is especially true when interval and density evaluation methods are applied (van Dijk et al., 2003). Hence, the results of Sarantis and Rapach and Wohar may well be attributed to the low power of out-of-sample predictability tests.

In this chapter, we attempt to shed light on the forecast performance of nonlinear real exchange rate models with respect to the linear AR and the RW benchmarks. We depart from the approach of previous studies and employ long spans of annual data for the dollar-sterling real exchange rate. By doing so, we extend the out-of-sample period to the entire post-Bretton Woods era. To our knowledge the forecasting performance of nonlinear real exchange rate models using long spans of data has not been examined so far.

Our modelling cycle consists of a battery of recently developed unit root tests, linearity tests, as well as bootstrap methods, which enable us to obtain a parsimonious specification of the nonlinear real exchange rate model. Subsequently, we employ the chosen specification and use Monte Carlo simulation techniques so as examine the empirical size and power properties of several forecast accuracy and encompassing tests.

Namely, we employ the t-type tests of Diebold and Mariano (1995), van Dijk and Franses (2003) and Harvey et al. (1998) as well as the F-type tests of Clark and McCracken (2005a). Our results indicate that all tests, with the exception of the test proposed by van Dijk and Franses (2003), have good size properties. This is a particularly important finding given the fact that the properties of F-type tests have not been examined when one of the competing models is nonlinear or nonstationary. Furthermore, we show that F-type tests have similar or substantially better power properties than their t-type counterparts. Unfortunately, both appear to exhibit low power for the comparison of nonlinear with linear AR models. Notwithstanding the above, our findings suggest that for the actual data the ESTAR model outperforms both the RW and AR benchmarks at short horizons for the majority of tests.

The rest of the chapter is structured as follows. Section 3.2 sets forth the STAR model and provides a description of the specification strategy adopted. The next section deals with generating forecasts from nonlinear models. Section 3.4 describes the forecast evaluation measures employed as well as the parametric bootstrap methodology for conducting statistical inference. Section 3.5 describes the empirical results for the actual real exchange rate data and the simulation exercise. The final section concludes.

3.2 Smooth Transition Models

The basic STAR model representation for a univariate time series $\{y_t\}$ is given by

$$\begin{aligned} y_t \; = \; & \pi_{1,0} + \pi_{1,1} y_{t-1} + \cdots + \pi_{1,p} y_{t-p} + (\pi_{2,0} + \\ & + \pi_{2,1} y_{t-1} + \cdots + \pi_{2,p} y_{t-p}) F(y_{t-1}; \gamma, c) + \epsilon_t, \quad t = 1, \ldots, T, \end{aligned} \tag{3.1}$$

or equivalently

$$y_t = \boldsymbol{\pi}_1' \boldsymbol{x}_t + \boldsymbol{\pi}_2' \boldsymbol{x}_t F(y_{t-1}; \gamma, c) + \epsilon_t, \qquad t = 1, \ldots, T, \tag{3.2}$$

where $\boldsymbol{x}_t = (1, \tilde{\boldsymbol{x}}_t')'$ with $\tilde{\boldsymbol{x}}_t = (y_{t-1}, \ldots, y_{t-p})'$, and $\boldsymbol{\pi}_j = (\pi_{j,0}, \ldots, \pi_{j,p})'$ for $j = 1, 2$. It is assumed that the error term, ϵ_t, is a martingale difference sequence. That is, $E[\epsilon_t | \mathcal{I}_{t-1}] = 0$, where $\mathcal{I}_{t-1}$ is the information set up to time $t-1$ consisting of all lagged values of y. The transition variable is given by the lagged endogenous variable y_{t-1} and c is a constant. The function $F(\cdot)$ is at least fourth-order, continuously differentiable with respect to the transition (or smoothness) parameter γ.

There are two common forms of the STAR model. The one we will discuss here in detail is the Exponential STAR (ESTAR) model of Teräsvirta (1994), in which transitions between a continuum of regimes are assumed to occur smoothly and symmetrically. The transition function $F(\cdot)$ of the ESTAR model is

$$F(y_{t-1};\gamma,c) = [1 - \exp(-\gamma(y_{t-1} - c)^2)]. \tag{3.3}$$

This transition function is symmetric around $(y_{t-1} - c)$ and admits the limits,

$$\begin{aligned} F(\cdot;\gamma,c) &\rightarrow 1 \text{ as } |y_{t-1} - c| \rightarrow +\infty, \\ F(\cdot;\gamma,c) &\rightarrow 0 \text{ as } |y_{t-1} - c| \rightarrow 0. \end{aligned}$$

Parameter γ can be seen as the transition speed of the function $F(\cdot)$ towards 1 (0) as the deviation grows larger (smaller). We are particularly interested in the special case that there is a unit root in the linear polynomial, $\sum_{i=1}^{p} \pi_{1,i} = 1$, $\pi_{2,i} = -\pi_{1,i}$ $\forall\ i \geq 1$, $\pi_{1,0} = 0$ and $c = \pi_{2,0}$. Under these restrictions, Equation (3.1) becomes

$$\begin{aligned} y_t &= \pi_{2,0} + [\pi_{1,1}(y_{t-1} - \pi_{2,0}) + \cdots + \pi_{1,p}(y_{t-p} - \pi_{2,0})] \\ &\quad \times \exp(-\gamma(y_{t-1} - \pi_{2,0})^2) + \epsilon_t. \end{aligned} \tag{3.4}$$

The above formulation is very appealing for modelling real exchange rates (see, e.g, Kilian and Taylor, 2003; Paya et al., 2003). Unlike in a linear model, the process moves between a white noise and a unit root depending on the size of the deviation from PPP, $|y_{t-1} - \pi_{2,0}|$. This type of adjustment is in accordance with the implications of theoretical models, which demonstate how frictions in international trade can induce nonlinear but mean reverting adjustment of the real exchange rate (see, e.g., Dumas, 1992; Berka, 2005). The rational is that small deviations are left uncorrected since they do not to cover transactions costs or the sunk costs of international arbitrage. On the other hand, large deviations are much less persistent. Therefore, the process exhibits strong persistence and near unit root behaviour.

Although ESTAR models can parsimoniously capture the adjustment mechanism proposed by theoretical models, their superiority in forecasting over rival models, like the RW and the AR, is clearly regime dependent. For instance, at the equilibrium, the process behaves similar to the RW, which implies that one cannot extract forecasting gains from using ESTAR models. On the other hand, substantially better forecasts can be obtained when large absolute deviations occur and the process is mean reverting fast. For AR models the speed of mean reversion is independent from the size of PPP deviations which results in substantial underestimation (overestimation) of the speed of mean reversion only for relatively large (small) deviations. The other common form of STAR models is the Logistic STAR (LSTAR), where the transition function $F(\cdot)$ is given by

$$F(\cdot;\gamma,c) = [1+\exp(-\gamma(y_{t-1}-c))]^{-1}.$$

The logistic transition function is asymmetric about $(y_{t-1} - c)$ and admits the limits,

$$\begin{aligned} F(\cdot;\gamma,c) &\rightarrow 1 \text{ as } (y_{t-1}-c) \rightarrow +\infty, \\ F(\cdot;\gamma,c) &\rightarrow 0 \text{ as } (y_{t-1}-c) \rightarrow -\infty. \end{aligned}$$

LSTAR models have also been fitted to real exchange rates (see Sarantis, 1999). Even though the theoretical argument is not as strongly supported as with the case of the ESTAR, there are some attempts to rationalise the asymmetric adjustment in the real exchange rate (see Campa and Goldberg, 2002).

We point out that as $\gamma \rightarrow 0$ the exponential and logistic transition functions approach a constant and both models collapse to a linear AR model. For the ES-

TAR model the same also holds when $\gamma \to \infty$.[2] The fact that STAR models nest linear AR models has important implications regarding the asymptotic distribution of commonly used forecast accuracy and encompassing tests (see, e.g., Clements and Galvão, 2004).

3.2.1 Linearity and Unit Root Tests

The fact that there is uncertainty about the exact DGP of real exchange rates motivates the use of data driven methods for the specification of parsimonious empirical models. In this study, we employ several testing procedures so as to examine whether the long-span real exchange rate series exhibits mean reversion and smooth transition dynamics. The rest of this section describes the linearity tests of Escribano and Jordá (1999) and Harvey and Leybourne (2007), and the unit root tests of Kapetanios et al. (2003) and Kapetanios and Shin (2008).[3]

Testing for the nonlinear part of Equation (2.2) gives rise to a nuisance parameter problem (Davies, 1977). Consequently, classical Lagrange Multiplier (LM) and Wald statistics may not follow standard distributions. In order to circumvent this problem, Luukkonen et al. (1988) suggest replacing the transition function by a Taylor series approximation around $\gamma = 0$. Escribano and Jordá (1999) build upon the work of Luukkonen et al. (1988) and propose the following auxiliary regression

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \boldsymbol{x}_t y_{t-1} + \boldsymbol{\delta}_2' \boldsymbol{x}_t y_{t-1}^2 + \boldsymbol{\delta}_3' \boldsymbol{x}_t y_{t-1}^3 + \boldsymbol{\delta}_4' \boldsymbol{x}_t y_{t-1}^4 + u_t \qquad (3.5)$$

for testing linearity and distinguishing between ESTAR and LSTAR processes. The null hypothesis of linearity corresponds to $H_0^1 : \boldsymbol{\delta}_1' = \boldsymbol{\delta}_2' = \boldsymbol{\delta}_3' = \boldsymbol{\delta}_4' = 0$ and the selection procedure between ESTAR and LSTAR is

1. Test the null of LSTAR nonlinearity, $H_0^L : \boldsymbol{\delta}_2' = \boldsymbol{\delta}_4' = 0$, with an F test,

[2] Moreover, the LSTAR model nests the TAR model since for $\gamma \to \infty$ the logistic transition function approaches the indicator function.

[3] For a more detailed discussion of linearity and unit root tests see (Pavlidis et al., 2009a) and Chapter 2.

(F_L).

2. Test the null of ESTAR nonlinearity, $H_0^E : \delta_1' = \delta_3' = 0$, with an F test, (F_E).

3. If the p-value of F_L is lower than F_E then select an ESTAR. Otherwise, select an LSTAR.

The use of the F-test is based on the assumptions that the process under examination is stationary and the error term in Equation (3.2) is i.i.d. However, a major concern in the PPP literature is that real exchange rates exhibits a unit root in which case the asymptotic distribution of linearity tests changes (Kiliç, 2004). Therefore, in order to avoid false inference one should first test for a unit root in the real exchange rate series. If the unit root hypothesis is rejected, the i.i.d. assumption can be relaxed by employing the wild bootstrap method (see Pavlidis et al., 2009b).

Harvey and Leybourne (2007) derive a more general linearity test statistic which has the same critical values under the null hypotheses of a linear I(0) and a linear I(1) processes. Rejection of the null therefore is indicative of nonlinearity and cannot be attributed to a linear I(1) DGP.

The Harvey and Leybourne test procedure consists of two steps. First is the test of linearity. Second, the order of integration of the linear or nonlinear process is determined. Consider the case of an I(0) process. By setting $p = 1$ and taking a second-order Taylor series expansion of Equation (3.1) around $\gamma = 0$ we obtain

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + u_t. \tag{3.6}$$

Whilst, in the case of an I(1) variable, the Taylor expansion yields

$$\Delta y_t = \varphi_0 \Delta y_{t-1} + \varphi_1 (\Delta y_{t-1})^2 + \varphi_1 (\Delta y_{t-1})^3 + \varepsilon_t.$$

In order to combine both possibilities, I(0) and I(1), Harvey and Leybourne (2007) propose the following regression model

$$
\begin{aligned}
y_t &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-1}^2 + \alpha_3 y_{t-1}^3 + \alpha_4 \Delta y_{t-1} + \alpha_5 (\Delta y_{t-1})^2 + \\
&\quad + \alpha_6 (\Delta y_{t-1})^3 + \eta_t. \qquad (3.7)
\end{aligned}
$$

In the presence of serial correlation, Equation (3.7) is augmented with lags of the first difference of the dependent variable. The null hypothesis of linearity is H_0 : $\alpha_2 = \alpha_3 = \alpha_5 = \alpha_6 = 0$ against the alternative hypothesis (nonlinearity) H_1 : at least one of $\alpha_2, \alpha_3, \alpha_5, \alpha_6$ is different from zero. The corresponding Wald statistic is

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

where the restricted residual sum of squares (RSS_1) comes from an OLS regression of y_t on a constant, y_{t-1}, and Δy_{t-1}. As Harvey and Leybourne point out, the distribution of W_T under the null differs depending on whether the process followed by y_t is I(0) or I(1). In order to make the limiting distribution of W_T homogeneous under the null, they multiply it with a correction that is the exponential of a weighted inverse of the absolute value of the Augmented Dickey Fuller (ADF) statistic,[4]

$$
W_T^* = \exp(-b\,|ADF_T|^{-1}) W_T. \qquad (3.8)
$$

An expression for the value of b is provided such that, for a given significance level, the critical value of W_T^* coincides with that from a $\chi^2(4)$. They also prove that, under H_1, W_T^* is consistent at the rate $O_p(T)$. The second step is to test whether the series is an I(0) or an I(1) process.

We note that pretesting for a unit root is also important in selecting forecasting models. Diebold and Kilian (2000) illustrate that the conventional view of employing models in first-differences when the series under examination is highly persistent can lead to less accurate forecasts. To this end, the authors advocate the

[4]This approach is suggested by Vogelsang (1998).

application of unit root tests for choosing between levels and differences.

Kapetanios et al. (2003) develop a test of a unit root null against the alternative of a globally stationary ESTAR. Their test is also based on a Taylor approximation of the nonlinear autoregressive model. For simplicity, assuming $p = 1,$, $\pi_{1,1} = 1$, $\pi_{2,1} = -\pi_{1,1}$, and $c = 0$, then (3.1) becomes

$$y_t = y_{t-1} + \left[1 - \exp\left(-\gamma y_{t-1}^2\right)\right]\left(-y_{t-1}\right) + u_t. \tag{3.9}$$

Using the first-order Taylor expansion and rearranging yields

$$\Delta y_t = \delta y_{t-1}^3 + u_t. \tag{3.10}$$

Hence, the null and alternative hypotheses are $H_0 : \delta = 0$ and $H_1 : \delta < 0$, respectively. The corresponding t-statistic is given by

$$t_{\text{NL}} = \frac{\hat{\delta}}{\text{s.e.}(\hat{\delta})}, \tag{3.11}$$

where $\text{s.e.}(\hat{\delta})$ denotes the standard error of $\hat{\delta}$. The asymptotic distribution of t_{NL} converges weakly to a functional of Brownian motions.

The issue of possible residual autocorrelation can be addressed by augmenting Equation (3.10) with lags of the dependent variable. Further, in the presence of deterministic components, the authors suggest replacing y_t in Equation (3.10) with the residuals from the regression of $\boldsymbol{y}$ on an intercept (demean case) or an intercept and a time trend (detrend case).

Kapetanios and Shin (2008) proceed in the spirit of Elliott et al. (1996) by employing a GLS procedure in order to increase the power of the nonlinear unit root test. In the case of a mean and a time trend in the data, the first step of the testing procedure includes computing the GLS estimate of $\boldsymbol{\theta}$ in

$$y_t = \tilde{\boldsymbol{\theta}}' \boldsymbol{z}_t + \tilde{y}_t, \tag{3.12}$$

by regressing $\boldsymbol{y}_{\bar{\rho}} = (y_1, y_2 - \bar{\rho} y_1, \ldots, y_T - \bar{\rho} y_{T-1})'$ on $\boldsymbol{z}_{\bar{\rho}} = (\boldsymbol{z}_1, \boldsymbol{z}_2 - \bar{\rho}\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T - \bar{\rho}\boldsymbol{z}_{T-1})'$ where $\boldsymbol{z}_t = (1, t)'$ and $\bar{\rho} = 1 - \bar{c}/T$ so as to obtain the estimated residuals, $\tilde{y}_t$.[5] For the demean case $\boldsymbol{z}_t$ is replaced by $z_t = 1$. Subsequently, Equation (3.10) is fitted to the GLS demeaned or detrended series and the t-statistic, $t_{\text{NL}}^{\text{GLS}}$, corresponding to $H_0 : \delta = 0$ is obtained. Kapetanios and Shin (2008) illustrate that the $t_{\text{NL}}^{\text{GLS}}$ statistic, like the t_{NL}, has a non-standard distribution.

Researchers typically employ Heteroskedasticity Consistent Covariance Matrix Estimators in order to robustify unit root tests against heteroskedasticity of unknown form. Cook (2006) illustrates that in small samples this practice can lead to moderate oversizing of the ADF and the Kapetanios et al. (2003) tests. Pavlidis et al. (2007) draw a similar conclusion for the test of Kapetanios and Shin (2008). In order to address this issue we construct exact sample critical values for the heteroskedasticity-robust test statistics via stochastic simulation.

3.3 Forecasting with Nonlinear Models

A general dynamic model for the process $\{y_t\}$ can be written as

$$y_t = g(\boldsymbol{x}_t; \boldsymbol{\phi}) + \epsilon_t, \tag{3.13}$$

where $\boldsymbol{x}_t = (1, y_{t-1}, \ldots, y_{t-p})'$, $\boldsymbol{\phi}$ is a parameter vector and $\epsilon_t \sim \text{iid}(0, \sigma^2)$. By assuming a quadratic loss function the optimal h-step ahead forecast is $\hat{y}_{t+h|t} = E\left[y_{t+h} \mid \mathcal{I}_t\right]$, where $\mathcal{I}_t$ denotes the information set at time t (Clements, 2005). The complexity of generating forecasts from the above model depends crucially on function $g(\cdot)$. When $g(\cdot)$ is a linear operator so that Equation (3.13) specifies an AR(p), a closed-form solution always exist. In this context, one- and multi-step ahead forecasts can be easily obtained through recursion (see, e.g., Hamilton, 1994, ch. 4). On the other hand, when $g(\cdot)$ is a nonlinear function closed-form

[5] Kapetanios and Shin (2008) set $\bar{c}$ equal to -17.5 so that the asymptotic power of the test under the local alternative is 0.5.

solutions for multi-step forecasts are not generally available. In this case, the one-step ahead forecast is given by

$$\hat{y}_{t+1|t} \equiv E[y_{t+1} \mid \mathcal{I}_t] = E[g(\boldsymbol{x}_{t+1};\boldsymbol{\phi}) + \epsilon_{t+1}] = g(\boldsymbol{x}_{t+1};\boldsymbol{\phi}). \tag{3.14}$$

and, therefore, $\hat{y}_{t+1|t}$ can be computed in a analogous way to the linear AR(p) model. For larger forecast horizons, say $h = 2$, however,

$$\begin{aligned}
\hat{y}_{t+2|t} &\equiv E\left[y_{t+2} \mid \mathcal{I}_t\right] = E\left[g(\boldsymbol{x}_{t+2};\boldsymbol{\phi}) + \epsilon_{t+2} \mid \mathcal{I}_t\right] && (3.15)\\
&= E\left[g(\boldsymbol{x}_{t+2};\boldsymbol{\phi}) \mid \mathcal{I}_t\right] + E[\epsilon_{t+2}] && (3.16)\\
&= E\left[g(\boldsymbol{x}_{t+2};\boldsymbol{\phi}) \mid \mathcal{I}_t\right]. && (3.17)
\end{aligned}$$

Because the expected value of a nonlinear function is not necessarily equal to the function value calculated at the expectation of its argument, $E\left[g(\cdot)\right] \neq g\left(E[\cdot]\right)$,

$$\begin{aligned}
\hat{y}_{t+2|t} &\equiv E[g(\boldsymbol{x}_{t+2};\boldsymbol{\phi}) \mid \mathcal{I}_t] = E\left[g(\boldsymbol{x}_{t+2|t} + \epsilon_{t+2};\boldsymbol{\phi}) \mid \mathcal{I}_t\right] \neq && (3.18)\\
&\neq g\left(E[\boldsymbol{x}_{t+2} \mid \mathcal{I}_t] + E[\epsilon_{t+2} \mid \mathcal{I}_t];\boldsymbol{\phi}\right) = g(\hat{\boldsymbol{x}}_{t+2|t};\boldsymbol{\phi}). && (3.19)
\end{aligned}$$

It follows that a recursive relationship between forecasts at different horizons cannot be established.

A widely used method to approximate $E\left[y_{t+h} \mid \mathcal{I}_t\right]$ when $g(\cdot)$ is nonlinear is bootstrap integration.[6] The procedure of generating the h-step ahead forecast is

1. Use Equation (3.14) and $\mathcal{I}_t$ to compute $\hat{y}_{t+1|t}$.[7]

2. Randomly draw with replacement $h-1$ values from the estimated residuals $\hat{\epsilon}$ of the nonlinear time series model (3.13).

[6] Alternative methods proposed in the literature for constructing multi-step forecasts for nonlinear models are the Naive or Skeleton method, the Exact method and Monte Carlo simulation (e.g., Teräsvirta, 2006). The former method is based on recursive substitution by setting $E[g(\cdot)] = g(E[\cdot])$ and, therefore, produces biased results. The Exact method employs numerical integration, which requires assumptions regarding the error distribution and is computationally intensive for large forecast horizons. The final method, Monte Carlo simulation, is similar to the bootstrap but again requires distributional assumptions.

[7] Note that in practice there is parameter uncertainty since ϕ is not known and has to be estimated.

3. Use the bootstrap innovations, $\mathcal{I}_t$, and $\hat{y}_{t+1|t}$ obtained in the first step, and iterate on the nonlinear model so as to compute a forecast $\hat{y}_{t+h|t}$.

4. Repeat Steps 2 and 3 B times, where B is large number, so as to get $\hat{y}^i_{t+h|t}$, where $i = 1, \ldots, B$.

5. The h-step ahead boostrap forecast is given by

$$\hat{y}^{bs}_{t+h|t} = \frac{1}{B} \sum_{i=1}^{B} \hat{y}^i_{t+h|t}. \tag{3.20}$$

An attractive feature of the bootstrap method is that it does not require distributional assumption. The errors, however, are presumed to be iid. The results of Clements and Smith (1997) support the use of bootstrap methods in forecasting from nonlinear autoregressive models. For a survey on forecasting with STAR models see Lundbergh and Teräsvirta (2002).

3.4 Evaluating Forecasts

Forecast evaluation provides an alternative way of model selection. We restrict our attention to the comparison of point forecasts on the basis of forecast accuracy and forecast encompassing measures. The former measures include the MSE-t of Diebold and Mariano (1995), the MSE-F test of Clark and McCracken (2005a) and the Weighted MSE-t (W-MSE-t) proposed by van Dijk and Franses (2003). The latter are the Harvey et al. (1998) ENC-t and the ENC-F of Clark and McCracken (2005a).

Our setting is similar to the one adopted by Clark and McCracken (2005a). The number of in-sample and out-of-sample observations is denoted as R and P, respectively, so that the total number of observations is $T = R + P$. We adopt a recursive scheme for forecasting, where as t increases from R to $T - h$ the parameters of the models are re-estimated by employing data up to time t so as to generate forecast for the following h horizons. In accordance with the notation used in the previous section, y_{t+h} denotes the variable to be predicted at time $t = R, \ldots, T-h$ with the number of forecasts corresponding to horizon h being equal to $P-h+1$. The forecast errors are defined as $\hat{e}_{1,t+h} = y_{t+h} - \hat{y}_{1,t+h|t}$ for the benchmark model and $\hat{e}_{2,t+h} = y_{t+h} - \hat{y}_{2,t+h|t}$ for the competing model.

3.4.1 Tests of Forecast Accuracy

The first three tests examine forecast accuracy by setting the Mean Square Error (MSE) as the measure of predictive ability. In this setting, the null hypothesis is that the MSEs of the two competing models are equal against the one-sided alternative that the MSE for the second model is smaller. Diebold and Mariano (1995) develop the following widely used t-type test

$$\text{MSE} - t = (P-h+1)^{1/2}\frac{\bar{d}}{\widehat{S}_{dd}^{1/2}}, \tag{3.21}$$

where $\widehat{d}_{t+h} = \widehat{e}_{1,t+h}^2 - \widehat{e}_{2,t+h}^2$, $\bar{d} = (P-h+1)^{-1}\sum_{t=R}^{T-h}\widehat{d}_{t+h} = \text{MSE}_1 - \text{MSE}_2$, $\widehat{\Gamma}_{dd}(j) = (P-h+1)^{-1}\sum_{t=R+j}^{T-h}\widehat{d}_{t+h}\widehat{d}_{t+h-j}$ for $j \geqslant 0$ and $\widehat{\Gamma}_{dd}(j) = \widehat{\Gamma}_{dd}(-j)$, and $\widehat{S}_{dd} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\Gamma}_{dd}(j)$ denotes the long-run variance of d_{t+h} estimated using a kernel-based estimator with function $K(\cdot)$, bandwith parameter M and maximum number of lags $\bar{j}$.[8]

For non-nested models the long-run variance of $\hat{d}_{t+h}$ is positive and the MSE-t statistic follows asymptotically the standard normal distribution. When the number of forecasts is relatively small, Harvey et al. (1997) illustrate that a distinctive improvement of the test can be achieved by correcting for small-sample bias in the estimated variance of d_{t+h} and comparing the statistic to the Student's t distribution with $P-h$ degrees of freedom. The corrected test statistic is obtained by

[8]The use of Heteroskedasticity and Autocorrelation-Consistent (HAC) estimators for computing the variance of d_{t+h} is based on the fact that h-steps-ahead forecast errors will be serially correlated of order $h-1$. The performance of the MSE-t test using different HAC estimators is examined in Clark (1999).

multiplying MSE-t by

$$\zeta = \sqrt{\frac{P - 2h + h(h-1)/(P-h+1)}{(P-h+1)}}.$$

On the contrary, when the competing models are nested their population errors are identical under the null and, therefore, d_{t+h} and its variance are equal to zero. In this case, the asymptotic distribution of the statistic is non-standard and depends upon nuisance parameters for $h \geq 2$ (McCracken, 2004).[9]

The degeneracy of the long-run variance of d_{t+h} motivates Clark and McCracken (2005a) to propose a variant of the above test for nested models. Inspired by the in-sample F-test the author suggests replacing $\widehat{S}_{dd}^{1/2}$ with the variance of the forecast error of the "unrestricted" model. The new test statistic is given by

$$\text{MSE} - F = (P - h + 1)^{1/2} \frac{\bar{d}}{\text{MSE}_2}. \tag{3.22}$$

and has, asymptotically, better power properties (Clark and McCracken, 2005a). The limiting distribution of the MSE-F test statistic, like the MSE-t, is free of nuisance parameters only for $h = 1$ and is non-standard.

The forecast accuracy tests examined so far attach equal importance to all forecasts irrespectively of the available information set at time t. Hoverer, given the properties of the nonlinear adjustment mechanism for the real exchange rate, a researcher would expect the superiority of the ESTAR model over the RW to become most apparent for large deviations of the process from its equilibrium value. While for smaller deviations the two models should perform similarly. van Dijk and Franses (2003) propose a forecast evaluation test that employes a weighted average loss differential and comprises a modification of the MSE-t of

[9]The asymptotic distributions of all the test statistics for multi-step forecasts from nested models under parameter uncertainty are derived in Clark and McCracken (2005a). However, their derivation is based on the sufficient but not necessary assumptions of stationarity and linearity of the parameters, which are clearly not satisfied in our experiment.

Diebold and Mariano (1995). The corresponding test statistic is

$$\mathrm{W-MSE}-t=(P-h+1)^{1/2}\frac{\bar{d}^w}{\widehat{S}^{1/2}_{d^w d^w}}, \tag{3.23}$$

where $\hat{d}^w_{t+h} = w(\boldsymbol{x}_t) \times (\hat{e}^2_{1,t+h} - \hat{e}^2_{2,t+h})$, $\bar{d}^w = (P-h+1)^{-1}\sum_{t=R}^{T-h}\hat{d}^w_{t+h}$, $\widehat{\Gamma}_{d^w d^w}(j) = (P-h+1)^{-1}\sum_{t=R+j}^{T-h}\hat{d}^w_{t+h}\hat{d}^w_{t+h-j}$ for $j \geqslant 0$ and $\widehat{\Gamma}_{d^w d^w}(j) = \widehat{\Gamma}_{d^w d^w}(-j)$, and $\widehat{S}_{d^w d^w} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\Gamma}_{d^w d^w}(j)$ denotes the long-run variance of d^w_{t+h} estimated using a kernel-based estimator with function $K(\cdot)$, bandwith parameter M and maximum number of lags $\bar{j}$. The weight function is given by

$$w(\boldsymbol{x}_t) = 1 - \frac{f(y_t)}{\max(f(y_t))}, \tag{3.24}$$

where $f(\cdot)$ is the density function of y_t, so that more importance is attached to forecasts corresponding to deviations at the tails of the distribution. van Dijk and Franses (2003) show that the modified test statistic follows the same distribution with the MSE-t.

3.4.2 Forecast Encompassing

The remaining tests concern forecast encompassing. Consider the following combination of forecasts from the two competing models

$$y_{c,t+h|t} = (1-\lambda)y_{1,t+h|t} + \lambda y_{2,t+h|t}, \tag{3.25}$$

where $\lambda \in [0,1]$. Letting $e_{c,t+h}$ denote the error of the composite forecast $y_{c,t+h|t}$ and substituting $y_{1,t+h|t}$ and $y_{2,t+h|t}$ yields

$$e_{1,t+h|t} = \lambda(e_{1,t+h|t} - e_{2,t+h|t}) + e_{c,t+h|t}. \tag{3.26}$$

In this case, the null hypothesis that the forecast of the benchmark model incorporates all the relevant information in the forecast of the competing model is given by

$H_0 : \lambda = 0$. That is, the covariance between the forecast errors of the first model and the difference of the forecasts errors of the two models is equal to zero (see West, 2006). Under the alternative, the covariance is positive, $H_1 : \lambda > 0$, indicating that the second model has additional predictive power. Clearly, the forecast encompassing tests are also one-sided to the right. Harvey et al. (1998), based on the work of of Diebold and Mariano (1995), derive the following forecast-encompassing test statistic[10]

$$\mathrm{ENC} - t = (P - h + 1)^{1/2} \frac{\bar{c}}{\widehat{S}_{cc}^{1/2}}. \tag{3.27}$$

Let $\widehat{c}_{t+h} = \widehat{e}_{1,t+h}(\widehat{e}_{1,t+h} - \widehat{e}_{2,t+h})$, $\widehat{\Gamma}_{cc}(j) = (P - h + 1)^{-1} \sum_{t=R+j}^{T-h} \widehat{c}_{t+h}\widehat{c}_{t+h-j}$ for $j \geqslant 0$ and $\widehat{\Gamma}_{cc}(j) = \widehat{\Gamma}_{cc}(-j)$, and let $\widehat{S}_{cc} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\Gamma}_{cc}(j)$ denote the long-run variance of c_{t+h}.

Clark and McCracken (2001) illustrate that the distribution of the ENC-t statistic converges to the same type of distribution with the MSE-t statistic when the forecasts are generated from linear nested models. By employing the same reasoning with the one used for the ENC-F test they propose the following F-type test statistic

$$\mathrm{ENC} - F = (P - h + 1)^{1/2} \frac{\bar{c}}{MSE_2}, \tag{3.28}$$

which again has a non-standard limiting distribution and depends on nuisance parameters for $h \geq 2$. Similarly to forecast accuracy measures, the F-type test has, asymptotically, greater power than its t-type counterpart.

Due to the fact that standard distribution theory may not apply, we conduct statistical inference by employing a parametric bootstrap method similar to Kilian (1999) and Kilian and Taylor (2003). The simulation exercise consist of the following steps

1. Employ the original real exchange rate series and compute the above forecast evaluation measures for all forecast horizons.

[10] The authors employ the small sample correction of Harvey et al. (1997) for the MSE-t statistic.

2. Estimate the restricted model for the real exchange rate (the RW or the AR model) using the whole sample in order to obtain the fitted residuals and coefficients.

3. Set the estimated model as the Null DGP and randomly draw with replacement from the residuals so as to create an artificial series for the real exchange rate with the same length as the actual series. Initialise the process by employing the observed values of the series.

4. Repeat the forecasting exercise using the artificial data so as to compute h bootstrap test statistics for each forecast evaluation measure.

5. Repeat steps 3 and 4 B times, where B is a large number, so as to obtain the bootstrap distributions of the test statistics under the null.

6. Compute the bootstrap p-value as the percentage of times the simulated statistic is more extreme than the original statistic.

7. Reject the null if the p-value is smaller than the chosen significance level.

Clark and McCracken (2005a) illustrate that when forecasts are generated from linear nested models this method performs adequately in terms of size and power even when the bootstrap model is not properly specified. However, the performance of the bootstrap technique as well as the validity of the F-type tests has not been explored when one of the competing models is nonlinear or the process is nonstationary. We contribute to the literature on nonlinear real exchange rate forecasting by examining the finite properties of F-type tests as well as their implications in the following section.

3.5 Empirical Results

We extend the dollar-sterling real exchange rate data set of Lothian and Taylor (1996) by using annual data for the U.S. and U.K. consumer price indices and the dollar-sterling nominal exchange rate obtained from the International Financial Statistics database. The updated series covers the period from 1791 to 2005 and is illustrated in Figure 3.1. The number of in-sample observations, R, is set equal to 183 which corresponds to the pre-Bretton Woods era (1791-1973) and the remaining 32 years, P, comprise the out-of-sample period.

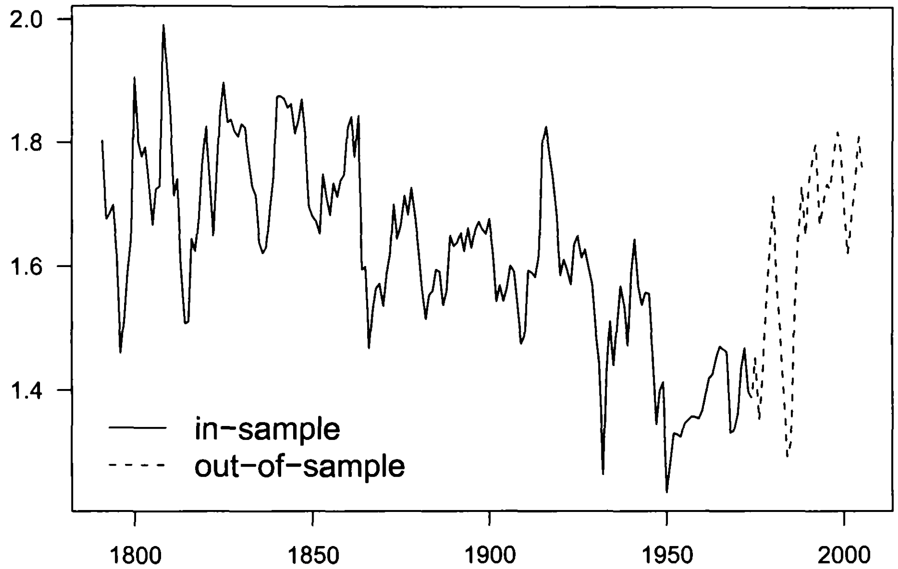

Figure 3.1: Time series plot of the dollar-sterling real exchange rate. The solid (dashed) line represents the in-sample (out-of-sample) period.

3.5.1 In-Sample Tests

Starting with the in-sample tests, we present results for both the entire sample period and the subperiod from 1791 to 1973. Table 3.1 reports the ADF, t_{NL} and $t_{\mathrm{NL}}^{\mathrm{GLS}}$ tests statistics as well as their heteroskedasticity-robust versions (ADF-HC, t_{NL}-HC, $t_{\mathrm{NL}}^{\mathrm{GLS}}$-HC) corresponding to the demean and detrend cases.[11] Exact sample critical values are constructed via stochastic simulation. For the demeaned real exchange rate, the unit root hypothesis is rejected by all tests at the 5% significance level. The only exception is the heteroskedasticity-robust version of the Kapetanios and Shin (2008) test, which rejects the null only at the 10% when data

[11]The lag length for the unit root and linearity tests is set to two on the basis of the Akaike Information Criterion.

prior to the recent floating period are used. Turning to the detrend case, the number of rejection decreases with the t_{NL}-HC statistic for the subperiod 1791-1973 and the $t_{\text{NL}}^{\text{GLS}}$ and $t_{\text{NL}}^{\text{GLS}}$-HC tests statistics for the whole period being larger than the corresponding 10% critical values. The mean reverting behaviour of the long-span real exchange rate is consistent with the empirical literature on PPP (see Frankel, 1990; Lothian and Taylor, 1996). Given the stationarity of the series, we follow the recommendation of Diebold and Kilian (2000) and choose to work with levels rather than first differences.

Table 3.1: Unit Root Tests

Sample Period: 1791-1973						
Case	ADF	ADF-HC	t_{NL}	t_{NL}-HC	$t_{\text{NL}}^{\text{GLS}}$	$t_{\text{NL}}^{\text{GLS}}$-HC
Mean	-3.082^{**}	-3.321^{**}	-3.488^{**}	-4.687^{**}	-2.211^{*}	-2.866^{**}
Trend	-4.985^{**}	-5.192^{**}	-3.707^{**}	-3.469	-3.824^{**}	-3.648^{**}
Sample Period: 1791-2005						
Case	ADF	ADF-HC	t_{NL}	t_{NL}-HC	$t_{\text{NL}}^{\text{GLS}}$	$t_{\text{NL}}^{\text{GLS}}$-HC
Mean	-3.794^{**}	-3.991^{**}	-4.522^{**}	-5.314^{**}	-2.873^{**}	-3.258^{**}
Trend	-4.327^{**}	-4.532^{**}	-4.406^{**}	-5.013^{**}	-2.293	-2.598

NOTE: ADF, t_{NL} and t_{NL} are the Augmented Dickey Fuller, the Kapetanios et al. (2003) and the Kapetanios and Shin (2008) unit root tests statistics. HC indicates heteroskedasticity-robust versions. $**$ and $*$ denote significance at the 5%, and 10% significance level, respectively. Critical values are constructed via Monte Carlo simulations.

We proceed by examining the presence of STAR-type nonlinearities by applying the Escribano and Jordá (1999) and Harvey and Leybourne (2007) testing procedures. The results are reported in Table 3.2. First, the wild bootstrap p-values for the tests developed in Pavlidis et al. (2009b) (top panel) corresponding to the null of linearity is marginally lower than the 5% significance level for the whole sample and slightly higher than the 10% for the subsample. Second, the fact that the p-value corresponding to F_L is lower than the one corresponding to F_E favours the use of the ESTAR model over the asymmetric LSTAR. Moreover, the Harvey and Leybourne (2007) test statistic is also greater than the 10% critical value which provides further support for the smooth transition model. The magnitudes of the p-values corresponding to the linearity tests indicate that the nonlinear mean-reverting behaviour of the series is more evident for the whole sample period than the pre-Bretton Woods era. This finding can be attributed to the higher power of the tests for larger sample sizes.

Table 3.2: Linearity Tests

	Escribano and Jordá (1999)		
Period	F	F_L	F_E
1791-1973	1.192 (0.114)	0.458 (0.610)	0.368 (0.695)
1791-2005	1.582 (0.043)	1.050 (0.244)	0.886 (0.329)

	Harvey and Leybourne (2007)
Period	W_T^*
1791-1973	8.494 (0.078)
1791-2005	10.478 (0.033)

NOTE: p-values are reported in parentheses. For the Escribano and Jordá (1999) test p-values are obtained through the wild bootstrap procedure described in Pavlidis et al. (2009b).

Next, we follow Kilian and Taylor (2003) and model the level of the real exchange rate using the ESTAR parameterisation (3.4).[12] Table 3.3 shows the estimates of the ESTAR model for the two periods examined, the standard error of the regressions, the corresponding t-statistics, the Ljung-Box Q-statistics for serial correlation in the residuals and the LM test statistic (ARCH) for conditional heteroskedasticity up to lags 1 and 5, and the wild bootstrap p-value for the transition parameter $\hat{\gamma}$. The Q and ARCH statistics do not indicate the presence of serial correlation or ARCH effects in the regression residuals. Moreover, the p-value is virtually zero in both cases suggesting that the estimated transition parameters are significant at all conventional levels. In accordance with the linearity tests results,

[12]Equation (3.4) imposes that the autoregressive coefficients sum to unity so that the process has a unit root in the inner regime. We test this restriction by running a Wald F-test. The corresponding p-value is substantially larger than 10% implying that the restricted version is also supported by the data.

the p-value for the transition parameter is lower for the whole sample illustrating that the degree of nonlinearity is more pronounced when longer spans of data are examined.

Table 3.3: Estimated Nonlinear Real Exchange Rate Model

Sample Period: 1791-1973

$$\hat{y}_t - \underset{(63.598)}{1.586} = (\underset{(13.834)}{1.122}(y_{t-1} - \underset{(63.598)}{1.586}) + (1 - \underset{(13.834)}{1.122})(y_{t-2} - \underset{(63.598)}{1.586}))$$
$$\times \exp(-\underset{\substack{(3.508)\\ [0.005]}}{2.076}(y_{t-1} - \underset{(63.598)}{1.586})^2).$$

$s = 0.067$; $Q_1 = 0.005$ [0.942]; $Q_5 = 3.941$ [0.558]; $\text{ARCH}_1 = 0.059$ [0.809]; $\text{ARCH}_5 = 0.220$ [0.953].

Sample Period: 1791-2005

$$\hat{y}_t - \underset{(81.518)}{1.590} = (\underset{(16.053)}{1.185}(y_{t-1} - \underset{(81.518)}{1.590}) + (1 - \underset{(16.053)}{1.185})(y_{t-2} - \underset{(81.518)}{1.590}))$$
$$\times \exp(-\underset{\substack{(4.357)\\ [0.000]}}{2.504}(y_{t-1} - \underset{(81.518)}{1.590})^2).$$

$s = 0.068$; $Q_1 = 0.002$ [0.963]; $Q_5 = 4.133$ [0.530]; $\text{ARCH}_1 = 0.079$ [0.778]; $\text{ARCH}_5 = 0.416$ [0.837].

NOTE: Figures in parentheses and square brackets denote absolute t-statistics and p-values, respectively. The p-value for the transition parameter $\hat{\gamma}$ is obtained through a simulation exercise, where the bootstrap DGP is the unit root model. s is the standard error of the regression. Q_1 and Q_5 denote the Ljung-Box Q-statistic for serial correlation up to order 1 and 5, respectively. ARCH_1 and ARCH_5 denote the LM test statistic for conditional heteroskedasticity up to order 1 and 5, respectively.

3.5.2 Out-of-Sample Tests

The in-sample test results provide strong support for a nonlinear adjustment mechanism of the real exchange rate. We now turn to the forecasting exercise. As we highlighted in the previous sections: (i) out-of-sample tests are likely to exhibit lower power than in-sample tests, and (ii) there is uncertainty regarding the validity of the F-type tests when one of the competing model is nonlinear or nonstationary. These motivate us to examine the small sample properties of the forecast evaluation measures by conducting a set of Monte Carlo simulation experiments. The nominal significance level is set equal to 5% for all experiments, the maximum forecast horizon equal to 4 and the number of bootstrap replication, B, equal to 1,000.

Empirical Size of Forecast Evaluation Tests

Initially, we focus on the empirical size of the tests, which is computed by the following procedure

1. Fit the benchmark model (the RW or the linear AR) to the whole sample.

2. Set the fitted model as the Null DGP and generate 1,000 artificial series of size equal to the size of the actual real exchange rate series.[13]

3. For each series adopt the same setting as for the actual data and generate forecasts from the benchmark and the competing model(s).

4. Apply the bootstrap methodology outlined in Section 3.4 so as to compute a vector of bootstrap p-values.

5. The empirical size of the test is defined as the percentage of times the bootstrap p-value is smaller than the 5 % significance level.

[13]Fake series are generated by drawing from the normal distribution with variance equal to the variance of the actual residuals. The first observations of the actual data are employed as initial values.

The results for the case of the RW against the ESTAR (RW-ESTAR), the RW against the AR (RW-AR) and the AR against the ESTAR (AR-ESTAR) are presented in Table 3.4. A broad conclusion that emerges is that the empirical size of all tests, but the W-MSE-t, is close to the nominal level with no test consistently outperforming the others.

Table 3.4: Empirical Size of Forecast Evaluation Tests

RW-ESTAR					
Horizon	MSE-t	W-MSE-t	ENC-t	MSE-F	ENC-F
1	0.056	0.089	0.061	0.058	0.058
2	0.058	0.079	0.053	0.056	0.048
3	0.054	0.072	0.056	0.055	0.047
4	0.038	0.056	0.039	0.058	0.045
RW-AR					
Horizon	MSE-t	W-MSE-t	ENC-t	MSE-F	ENC-F
1	0.055	0.104	0.055	0.052	0.051
2	0.046	0.087	0.044	0.045	0.042
3	0.046	0.071	0.040	0.051	0.044
4	0.041	0.063	0.041	0.053	0.041
AR-ESTAR					
Horizon	MSE-t	W-MSE-t	ENC-t	MSE-F	ENC-F
1	0.043	0.034	0.040	0.067	0.066
2	0.046	0.023	0.047	0.050	0.055
3	0.049	0.032	0.051	0.056	0.054
4	0.052	0.033	0.052	0.059	0.047

NOTE: The table shows the empirical size of the MSE-t, W-MSE-t, ENC-t, MSE-F and ENC-F test statistics for the RW-ESTAR, RW-AR and AR-ESTAR pairs. The nominal significance level is 5% and the horizons considered are $h = 1, \ldots, 4$.

The (absolute) error in rejection probability reaches a maximum of just 1.7 percentage points (for the MSE-F at the one year horizon). Most importantly, these results indicate that F-type tests are valid in our nonlinear context. As far as the W-MSE-t is concerned, the test exhibits moderate size distortions of up to 5 percentage points. Specifically, for the RW-ESTAR and the RW-AR cases the test is oversized at short horizons with the empirical size taking values close to 10%. While, for the AR-ESTAR case the weighted MSE-t statistic becomes undersized with the empirical size reaching a minimum value equal to 0.023 at $h = 2$.

Empirical Power of Forecast Evaluation Tests

We turn to the empirical power of the tests. The procedure adopted is identical to that for the size with the exception that the Null DGP is given by the estimated ESTAR model. Table 3.5 shows the results for the RW-ESTAR and AR-ESTAR cases. Overall, we observe that despite the fact that there are major differences across tests and pairs of competing models, the empirical power of all tests tends to decrease with the forecast horizon. Starting with the RW-ESTAR, t-type tests perform substantially worse than F-type tests. Specifically, the MSE-t ranks last with the empirical power ranging from about 15% for $h = 1$ to about 8% for $h = 4$. The W-MSE-t and ENC-t tests follow with the latter being marginally superior than the former but again with very low empirical power.[14] An increase by a factor of two or greater (depending on the horizon) in the frequency of rejecting the null occurs as we move to the MSE-F. The empirical power of the test exceeds 50%. Finally, the ENC-F test exhibits the highest power, which ranges from 68 to about 75%.

[14]The results for the W-MSE-t test should be interpreted with caution due to the poor size properties of the test.

Table 3.5: Empirical Power of Forecast Evaluation Tests

RW-ESTAR					
Horizon	MSE-t	W-MSE-t	ENC-t	MSE-F	ENC-F
1	0.152	0.170	0.239	0.577	0.752
2	0.094	0.124	0.133	0.588	0.730
3	0.078	0.096	0.111	0.566	0.709
4	0.079	0.095	0.114	0.528	0.680
AR-ESTAR					
Horizon	MSE-t	W-MSE-t	ENC-t	MSE-F	ENC-F
1	0.237	0.163	0.203	0.269	0.220
2	0.209	0.121	0.170	0.176	0.124
3	0.163	0.103	0.142	0.106	0.064
4	0.137	0.071	0.108	0.060	0.025

NOTE: The table shows the empirical power of the MSE-t, W-MSE-t, ENC-t, MSE-F and ENC-F test statistics for the RW-ESTAR and AR-ESTAR pairs. The nominal significance level is 5% and the horizons considered are $h = 1, \ldots, 4$.

As far as the AR-ESTAR pair is concerned, the performance of the F-type tests deteriorates while t-type tests exhibit similar empirical power to the RW-ESTAR case. The maximum power, which is achieved at $h = 1$ in all cases, ranges from about 16 (W-MSE-t) to about 27% (MSE-F). In other words, there is a small likelihood of identifying the forecasting gains from adopting an ESTAR rather than a linear AR model even though the true DGP process is nonlinear. These results are qualitatively similar with Clements and Smith (1999) for SETAR models. The low power of the tests suggests that superior in-sample but not out-of-sample performance of nonlinear models should not be documented as conclusive evidence against nonlinearity.[15]

[15]These results are similar to Inoue and Kilian (2005) regarding the comparison of linear mod-

Forecasting the Dollar-Sterling Real Exchange Rate

Table 3.6 presents the results regarding the comparison of forecasts for the actual real exchange rate series. The first three panels report the t-type test statistics, namely the MSE-t, W-MSE-t, and ENC-t tests. While the last two panels show the F-type tests statistics. The corresponding bootstrap p-values are reported in parentheses.

Table 3.6: Comparing Forecasts for the Dollar-Sterling Real Exchange Rate, 1974-2005

Panel A — MSE-t test			
Horizon	RW-ESTAR	RW-AR	AR-ESTAR
1	1.827 (0.047)	1.441 (0.091)	1.702 (0.019)
2	1.790 (0.067)	1.514 (0.129)	1.281 (0.048)
3	1.664 (0.114)	1.600 (0.134)	0.836 (0.098)
4	1.670 (0.118)	1.702 (0.123)	0.357 (0.203)
Panel B — W-MSE-t test			
Horizon	RW-ESTAR	RW-AR	AR-ESTAR
1	1.718 (0.053)	1.547 (0.066)	1.354 (0.032)
2	1.682 (0.069)	1.654 (0.083)	1.146 (0.062)
3	1.593 (0.095)	1.647 (0.099)	0.815 (0.121)
4	1.617 (0.097)	1.695 (0.104)	0.528 (0.190)
Panel C — ENC-t test			
Horizon	RW-ESTAR	RW-AR	AR-ESTAR
1	2.016 (0.066)	1.794 (0.105)	1.942 (0.033)
2	1.979 (0.105)	1.929 (0.134)	1.607 (0.062)

Continued on Next Page...

els.

Table 3.6: Comparing Forecasts (Continued)

3	1.942 (0.143)	2.077 (0.131)	1.157 (0.146)
4	2.054 (0.145)	2.276 (0.130)	0.641 (0.295)

Panel D — MSE-F test

Horizon	RW-ESTAR	RW-AR	AR-ESTAR
1	17.842 (0.000)	11.715 (0.002)	5.369 (0.046)
2	24.793 (0.000)	17.884 (0.008)	5.651 (0.108)
3	29.577 (0.002)	24.756 (0.016)	3.657 (0.181)
4	37.289 (0.007)	34.970 (0.017)	1.583 (0.254)

Panel E — ENC-F test

Horizon	RW-ESTAR	RW-AR	AR-ESTAR
1	22.449 (0.001)	15.616 (0.003)	6.361 (0.104)
2	30.060 (0.002)	23.754 (0.013)	6.934 (0.181)
3	36.531 (0.013)	32.791 (0.019)	4.729 (0.302)
4	47.439 (0.020)	46.750 (0.023)	2.528 (0.393)

NOTE: The table shows the MSE-t, W-MSE-t, ENC-t, MSE-F and ENC-F evaluation measures for the comparison of actual real exchange rate forecasts from the ESTAR, AR and RW models. Bootstrap p-values are reported in parentheses. The horizons considered are $h = 1, \ldots, 4$.

A broad conclusion that emerges is that as the forecast horizon increases the p-values for all tests tend to increase indicating that long-horizon predictability depends upon short-horizon predictability. This observation is consistent with the behaviour of the empirical power of the tests reported in the Table 3.5. Furthermore, the forecasting gains from using our nonlinear model specification are particularly evident at short forecast horizons. To this end, we mainly focus on one step ahead forecasts.

By examining the RW-ESTAR pair (second column), we observe that all five forecast encompassing and forecast accuracy test statistics are statistically significant at the 10% significance level. By changing the significance level to 5%, the null hypothesis is rejected by the two F-type tests and the MSE-t test (three out of the five cases). We note that for the F-type tests, p-values are close to zero for all forecasts horizons, which is not true for the t-type tests. The fact that F-type tests are associated with much lower p-values than their t-type counterparts when the benchmark model is the RW is not surprising given the higher empirical power of the former.

Turning to the RW-AR pair (third column), we generally observe higher p-values than for the RW-ESTAR pair. The number of rejections at the 10% level reduces from five to four for $h = 1$. While, at the 5% level only the two F-type tests reject the null.[16] Summarising the above results, both AR and ESTAR models appear to have predictive ability regarding the behaviour of the dollar-sterling real exchange rate.

The final column (AR-ESTAR) of Table 3.6 presents the results for the comparison of these two models. Despite the low empirical power of the forecast evaluation measures, at $h = 1$ all test statistics are significant at the 5% with the exception of the ENC-F, which has a p-value marginally higher than 10%. The number of rejections substantially reduces with the forecast horizon and at $h = 2$ only the MSE-t test rejects the null hypothesis. This may be due to the fact that both models share the prediction that the series will eventually mean revert to its equilibrium value.

Overall, the out-of-sample tests results complement those of the in-sample tests and provide strong support for the ESTAR model. In contrast to previous studies, which employ higher frequency data, our findings illustrate that nonlinear real exchange rate models are useful for forecasting the long-span real exchange

[16]Lothian and Taylor (1996) and Siddique and Sweeney (1998) also show that AR models provide superior forecasts (in terms of the RMSE criterion) to the RW for the recent float.

rate.

3.6 Conclusion

This chapter utilises a long span of data in order to investigate the ability of the ESTAR model to forecast the dollar-sterling real exchange rate. We pay special attention to model specification by employing several recently proposed linearity and unit root tests as well as bootstrap techniques. In turn, we investigate the small sample properties of a battery of forecast evaluation measures. Our results, in line with the literature on forecasting from nonlinear models, illustrate the difficulty of detecting the superiority of STAR models to AR models. Despite the low power of out-of-sample evaluation tests, we find that recursive ESTAR forecasts for the actual real exchange rate series outperform all rival forecasts. Consequently, researchers and practitioners can extract forecasting gains regarding the behaviour of the long-span real exchange rate series by employing nonlinear models.

CHAPTER 4

Further Empirical Evidence on the Consumption-Real Exchange Rate Anomaly

4.1 Introduction

In the early 1990s the lack of evidence supporting Purchasing Power Parity (PPP) led researchers to focus on the identification of potential pitfalls concerning the empirical approaches employed till then as well as to provide theoretical justifications for the observed behaviour of real exchange rates.

Three of the most important avenues of research that emerged have focused on: (i) the effect of the sample size, (ii) the presence of nonlinearities in the adjustment mechanism, and (iii) the fact that real variables may affect the equilibrium real exchange rate. As far as the latter factor is concerned International Real Business Cycle (IRBC) models, with complete or incomplete asset markets, establish a relationship between the equilibrium real exchange rate and consumption series on the basis of international risk sharing (e.g., Backus and Smith, 1993; Kollmann, 1995; Chari et al., 2002). However, the findings of a number of studies cast doubts on the empirical validity of this implication (see, e.g., Benigno and Thoenissen, 2008). The main objective of the present chapter is to reassess the implied relationship between the real exchange rate and consumption by extending the sample used by previous studies and by allowing the presence of nonlinearity in the adjustment mechanism. The rest of the introductory section outlines recent advances in the literature that motivate our approach.

As noted by Frankel (1986), the tests typically employed during the 1980s to investigate whether real exchange rates are stationary may have low power when applied to small spans of data during the recent floating rate period. Following Frankel a number of researchers supported this view by using long span of data (e.g., Lothian and Taylor, 1996) and panel unit root tests (Frankel and Rose, 1996). Even though these studies provided evidence that real exchange rates mean revert in the long-run, the implied half life of deviations from PPP ranged from three to five years. The fact that real shocks cannot account for such a high degree of persistence gave rise to Rogoff's (1996) PPP puzzle.

Perhaps the most important explanation of the Rogoff puzzle is provided by theoretical models which demonstrate how transactions costs or the sunk costs of international arbitrage induce nonlinear but mean reverting adjustment of the real exchange rate (see, e.g., Dumas, 1992; Sercu et al., 1995; O'Connell and Wei, 2002). Whilst globally mean reverting, these nonlinear processes have the property of exhibiting near unit root behaviour for small deviations from PPP, since small deviations are left uncorrected if they are not large enough to cover transactions costs or the sunk costs of international arbitrage, while large deviations are much less persistent. Hence, the low power of stationarity tests as well as the excess volatility of the real exchange rate may be attributed to the presence of nonlinearities in the data. In his seminal paper Dumas (1992) summarised this position as follows

"Linear equations are unlikely clearly to identify a process such as

the one for lnp[1] *... in which long-run behaviour is very different from short-term behaviour, since reversion manifests itself only when deviations from parity has become wide enough."*

Dumas (1992, p. 171)

The set of parametric models that can capture the nonlinearity postulated includes the Threshold Autoregressive (TAR) model of Tong (1983) and the Smooth Transition Autoregressive (STAR) model of Granger and Teräsvirta (1993) and Teräsvirta (1994). There are two common forms of the STAR model. The one is the Exponential STAR (ESTAR) model in which transitions between a continuum of regimes are assumed to occur smoothly and symmetrically. The appealing feature of the ESTAR model is that the speed of mean reversion is increasing with the size of the deviation from the equilibrium, which implies that the corresponding half life of a shock depends on its size. The smooth adjustment process is suggested in the analysis of Dumas (1992) and demonstrated by Berka (2002). Furthermore, Teräsvirta (1994) argues that if an aggregated process is observed, regime changes may be smooth rather than discrete as long as heterogeneous agents do not act simultaneously even if they individually make dichotomous decisions, which favours the use of the ESTAR model over TAR model.

Michael et al. (1997), Taylor et al. (2001) and Kilian and Taylor (2003) among others show that ESTAR models can parsimoniously fit a number of real exchange rates. Nonlinear impulse response functions derived from the estimated models suggest that large shocks mean revert much faster than the ones previously reported for linear models, for which the speed of mean reversion is independent of the size of the shock. These findings therefore seem to go some way towards solving Rogoff's PPP puzzle. However, deviations from PPP still dissipate very slowly.

[1] where lnp denotes the real exchange rate.

Although the early studies of PPP assumed a constant equilibrium rate, it is well recognised that even in relatively short spans of data real effects on the equilibrium exchange rate may be important. A variety of theoretical models, such as Balassa (1964) and Samuelson (1964), Lucas (1982) and Stein et al. (1995), demonstrate how real factors drive real exchange rates' movements and imply a non-constant equilibrium. Neglecting the influence of such factors may result in an omitted variable bias, which could account for the slow mean reversion reported in the empirical literature. The significance of real factors has been documented in panel data analysis (see, e.g., Canzoneri et al., 1996; Chinn and Johnston, 1996), as well as, in studies adopting a country by country nonlinear framework for long span of data (Lothian and Taylor, 2008; Paya and Peel, 2006a).

International Real Business Cycle (IRBC) models predict a close relation between movements in the real exchange rate and relative consumption levels. (e.g., Backus and Smith, 1993; Kollmann, 1995). However, the evidence in favour of a link between real exchange rate and relative consumption is scarce. Backus and Smith (1993) are the first to document the lack of a systematic pattern governing the movements of real exchange rates and relative consumption by comparing the means, standard deviations and autocorrelations of the first differences of the two series. Kollmann (1995) employs the methods proposed by Park (1992) and Phillips and Ouliaris (1990) to investigate whether consumption and real exchange rates are cointegrated. By using quarterly data for the recent floating period he concludes that the complete markets model cannot match the observed consumption and real exchange rate growth rates. This result also holds using panel data (Koedijk et al., 1996). More recently, Sercu and Uppal (2000) examine a different set of countries than the set used by Kollmann (1995) for the post-Bretton Woods era and find that there is a long-run relation between consumption and real exchange rates, on the basis of the Johansen (1991) test. However, the authors do not specify if the cointegration equation is consistent with the implications of IRBC models. Finally, Head et al. (2004) employ the GMM method and reject the hypothesis that there is a link between real exchange rates and relative consumption levels.

As noted by Obstfeld and Rogoff (2000) the fact that real exchange rates and consumption appear to be disconnected should be of no surprise given the high volatility of real exchange rates under floating together with the low volatility of consumption. The discrepancy between theory and empirical evidence is known as the "Backus and Smith puzzle" or the "consumption real exchange rate anomaly".

We argue that the empirical failure of IRBC models in previous studies may be due to the linear framework adopted in conjunction with the relatively short span of data available for the post-Bretton Woods era. Our line of reasoning is that factors such as the cost of arbitrage, the presence of heterogeneous agents (noise traders and rational speculators) in the market, and the fact that the equilibrium rate cannot be observed directly by the arbitrageurs may lead to persistent and inherently nonlinear deviations from economic fundamentals (e.g., Frankel and Froot, 1990; Kilian and Taylor, 2001; De Grauwe and Grimaldi, 2006). Moreover, we show that the sample correlation between the real exchange rate and relative consumption levels may be small or negative even though there is a well defined long-run structural relationship between the variables. Essentially the structural relationship is a nonlinear dynamic one so that the sample contemporaneous correlation may be misleading as to the structural relationship.[2]

When deviations from the equilibrium are small, arbitrageurs, who may be uncertain about the exact value of the equilibrium exchange rate, may be dominated by noise traders who can drive the exchange rate in the opposite direction. Hence, small misalignments of the exchange rate will be left uncorrected. However, when deviations from equilibrium become large a consensus is developed that the currency is overvalued or undervalued which, eventually, will result in driving the

[2]We are aware that real business cycle models that include incomplete asset markets , non-traded goods or other market frictions can explain the contemporaneous correlation (see, e.g., Chari et al., 2002; Kehoe and Perri, 2002; Benigno and Thoenissen, 2008; Selaive and Tuesta, 2006).

exchange rate towards its fundamental value. In this setting deviations from the equilibrium exhibit a high degree of persistence and smooth threshold dynamics. The hypothesis of a nonlinear adjustment to the equilibrium is also motivated by the empirical regularities noted by Backus and Smith (1993) and Obstfeld and Rogoff (2000), and with the difficulty of finding cointegration when linear models are used to analyse short spans of data.

The present study re-examines the validity of IRBC models during the recent floating period. By expanding the span of data used by previous studies we attempt to mitigate the low power of linear cointegration tests and to approximate the long-run relationship using the Johansen (1991) method. Subsequently, we apply the linearity test of Escribano and Jordá (1999) to the deviations from the IRBC equilibrium. We also consider two recent modifications of the linearity test which account for conditional heteroskedasticity. Our findings support the presence of smooth transition nonlinearity, which provides an explanation for the failure of cointegration tests based on relatively short span of data (e.g., Pippenger and Goering, 1993). It appears that STAR models produce parsimonious fits to the deviation series. The results of the Generalized Impulse Response Function (GIRF) suggest a fast adjustment process with half-lives between one to three years.

The rest of the chapter is structured as follows. Section 4.2 provides a brief discussion of IRBC models based on complete asset markets and ESTAR models. The next section describes the data, the empirical methodology and the experimental results. The final section concludes.

4.2 The Equilibrium Real Exchange Rate in IRBC models

International Real Business Cycle models comprise an extension of the closed economy Real Business Cycle models to an international setting where transactions take place both in goods as well as in financial markets (e.g., King et al., 1988). In this setting, as long as financial markets are complete, risk sharing takes place across countries with the real exchange rate being proportional to the ratio of marginal utilities of consumption (see, e.g., Chari et al., 2002; Apte et al., 2004). It follows that IRBC models with complete markets predict that higher real consumption abroad lowers the real value of foreign currency.

To analyse this statement more formally we follow Kollmann (1995) and assume a world with K countries indexed by $k = 1, \ldots, K$, each represented by an infinitely lived agent. Furthermore, it is assumed that the goods consumed differ across countries, which implies a non constant real exchange rate. Each country's preferences are given by

$$U_k = E_s \left[\sum_{t=s}^{\infty} \beta_k^{t-s} u_{k,t}(C_{k,t}) \right], \qquad k = 1, \ldots, K, \tag{4.1}$$

where, E is the expectations operator, $\beta_k \in (0,1)$ is country k's subjective discount factor, $u_{k,t}(\cdot)$ is country k's instantaneous utility function in period t, and $C_{k,t}$ denotes consumption of country k. In equilibrium, the risk sharing condition for any country pair (i, j) and for all periods and states is

$$Q_t = \Lambda_{i,j} \frac{\beta_i^t m_{i,t}}{\beta_j^t m_{j,t}}, \tag{4.2}$$

where Q_t is the real exchange rate[3] in period t, $\Lambda_{i,j}$ is a constant, and $m_{k,t}$ is the marginal utility of consumption for country $k = i, j$. The above relation should hold even if there are frictions in goods and labour markets, such as sticky prices, sticky wages, and shipping costs, because their effect is already reflected in consumption choices.

[3]The real exchange rate is defined as

$$Q_t = S_t \frac{P_{j,t}}{P_{i,t}}$$

where S_t denotes the nominal exchange rate, units of currency i per unit of currency j, and $P_{k,t}$ denotes consumer prices for country k.

Taking logs and assuming that the utility function is iso-elastic with exponent $1 - n_k$, where n_k denotes the coefficient of relative risk aversion of country $k = i, j$, Equation (4.2) yields the model tested by Kollmann (1995) and Backus and Smith (1993)[4]

$$q_t = \lambda_{i,j} + \ln\left(\frac{\beta_j}{\beta_i}\right) t + n_i c_{i,t} - n_j c_{j,t} + z_t, \tag{4.3}$$

where $q_t, \lambda_{i,j}, c_{i,t}$ and $c_{j,t}$ denote the logarithms of $Q_t, \Lambda_{i,j}, C_{i,t}$ and $C_{j,t}$, respectively, and z_t denotes the deviation from the equilibrium implied by the model. Given that the coefficient of risk aversion takes positive values, a country undergoing a real depreciation should experience relative consumption growth, with a rate depending on the elasticity of intertemporal substitution in consumption.

4.2.1 Nonlinear Adjustment to Equilibrium

Recently, a number of authors have provided evidence in favour of smooth nonlinear transition dynamics in the deviations of nominal exchange rates from macroeconomic fundamentals such as those suggested by the monetary model and the PPP (see, e.g., Taylor and Peel, 2000; Taylor et al., 2001; Paya and Peel, 2006a). A model that seems to parsimoniously capture the nonlinear mean reversion postulated is the ESTAR. An ESTAR model for the process $\{z_t\}$ may be written

$$z_t - \mu = \sum_{p=1}^{\bar{p}} \phi_p(z_{t-p} - \mu) \exp\left(-\gamma\left(z_{t-1} - \mu\right)^2\right) + \epsilon_t, \tag{4.4}$$

where $\gamma \in (0, \infty)$ is the smoothness parameter, which determines the transition speed of function $G(z_{t-1}; \gamma, \mu) = \exp\left(-\gamma\left(z_{t-1} - \mu\right)^2\right)$ towards the inner or outer regime. The error term, ϵ_t, is assumed to follow a white noise process with mean 0 and variance σ_ϵ, and μ is a constant. Equation (4.4) is a popular reformulation of the ESTAR model proposed by Granger and Teräsvirta (1993). The

[4]Backus and Smith (1993) derive a restricted model with identical risk aversion coefficients, as well as, subjective discount factors across countries. Whilst a more general model than the one of Kollmann (1995) is provided by Apte et al. (2004).

exponential transition function, $G(\cdot)$, is particularly applicable because it implies symmetric adjustment for positive and negative deviations from the equilibrium. Further, the speed of adjustment is increasing with the smoothness parameter γ and the absolute value of the past deviation from the equilibrium. A particularly interesting case is when $\sum_{p=1}^{\bar{p}} \phi_p = 1$. In this case, at the equilibrium $G(\cdot) = 1$ and z_t will behave as a unit root process, while for larger deviations $G(\cdot) \in [0, 1)$ and z_t will mean revert. Hence, although z_t is a globally mean reverting nonlinear process, it may exhibit a high degree of persistence, which provides an explanation for the low power of stationarity test. Kilian and Taylor (2003) propose a different ESTAR parameterisation. They argue that it is more intuitive to allow the effect of the deviations from the equilibrium on the nonlinear dynamics to be cumulative. To this end, the authors suggest modifying Equation (4.4) to

$$z_t - \mu = \sum_{p=1}^{\bar{p}} \phi_p (z_{t-p} - \mu) \exp\left(-\gamma \sum_{d=1}^{\bar{d}} (z_{t-d} - \mu)^2\right) + \epsilon_t, \tag{4.5}$$

where $\bar{d}$ is a positive integer. Suppose that $\bar{d}$ differs from unity and that the smoothness parameter, γ, is significant. Then cumulative deviations are a more informative indicator of whether the market is moving towards the equilibrium value rather than a single past deviation of the process.

4.3 Data, Empirical Methodology and Experimental Results

We use quarterly data for private consumption, nominal exchange rates and consumer price indices obtained from the International Financial Statistics database for Canada, Germany, France, Japan, Sweden, the United Kingdom and the United States. The sample period is from 1973:I to 2004:IV, except for Germany and France, for which the sample period ends at 1998:IV. We set the U.S. dollar as the reference currency for the empirical analysis.

In order to investigate whether the consumption real exchange rate anomaly is present in the examined data set, we initially utilise the correlation coefficients between real exchange rates and relative consumption. These correlations vary between -0.575 and 0.101,[5] indicating that the "Backus and Smith puzzle" remains for the extended sample period. However, the correlation statistic may be an inappropriate measure for testing the validity of IRBC models due to the presence of time trends in the equilibrium equation, different risk aversion parameters and nonlinear dynamics.

4.3.1 Cointegration Analysis

IRBC models clearly predict that there should be a long-run relationship between real exchange rates and consumption, or equivalently if the variables are integrated of order one, $I(1)$, they should form a cointegrating system. By the Granger Representation Theorem (Engle and Granger, 1987) the above set of variables must posses a Vector Error Correction Model (VECM) representation in which the error term, z_t, in Equation (4.3) comprises the deviations from the equilibrium. Let $\boldsymbol{y}_t = [q_t, c_{i,t}, c_{n,t}]$ denote the 3×1 vector of the system's variables, the VECM is written

$$\Delta \boldsymbol{y}_t = \sum_{i=1}^{p} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \tag{4.6}$$

where Δ is the difference operator. The rank of matrix $\boldsymbol{\Pi}$ determines the number of cointegrating relationships. If matrix $\boldsymbol{\Pi}$ is of full rank, $r = 3$, the VECM reduces to a vector autoregression (VAR) and $\boldsymbol{y}_t$ is a stationary process. If $\boldsymbol{\Pi}$ is the null matrix, $r = 0$, then the system's variables are not cointegrated and the underlying process is not stationary. Finally, if $\boldsymbol{\Pi}$ is neither of full rank nor the null matrix, $0 < r < 3$, then there are r cointegrating relationships and $\boldsymbol{\Pi}$ can be decomposed

$$\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{n}' \tag{4.7}$$

[5]These values are similar to the ones reported in the literature (see, e.g., Chari et al., 2002).

where n are the r cointegrating vectors determining the long-run equilibrium, and α denotes the matrix of the adjustment coefficients.

It is well recognised that depending on the properties of the series under examination cointegration techniques may have low power when applied to short spans of data. Further, due to serious small sample bias the coefficients obtained in cointegration analysis can vary widely across country pairs making economic interpretation very difficult (Froot and Rogoff, 1995). We examine this scenario by extending the data set used by previous studies and applying the Johansen (1991) methodology.

Table 4.1 reports the trace and λ-max (maximum eigenvalue) statistics for cointegration, and the long-run coefficients for consumption. Overall, the cointegration results support the existence of a long-run relationship among real exchange rates and consumptions. On the basis of both the trace and λ-max statistics the null of no cointegration can be rejected for all countries but Japan at the 10 % significance level.[6] Both tests indicate that there is a single cointegrating relationship between the system's variables. As the last two columns of Table 4.1 report, the long-run coefficients, n_{US} and n_j, are correctly signed for Canada, France, Sweden and the United Kingdom suggesting that higher (lower) real consumption abroad lowers (increases) the real value of foreign currency (see Equation (4.3)). The implied relative risk aversion parameters are sometimes higher than the upper limit of ten suggested as reasonable by Rajnish et al. (1985). However, recent work by Barro (2005) suggests that higher values may be realistic. It is noted that the correlation coefficients for these countries with the exception of France are negative. Therefore, the difference in the relative risk aversion coefficients and/or the presence of a time trend (due to different discount rates between countries) may result in negative contemporaneous correlations between the real exchange rate and relative consumption.

[6] Japan is excluded from the remaining analysis since no cointegration was found.

Table 4.1: Cointegration Results

			Trace			λ-max				
Country j	p	trend	$r <= 2$	$r <= 1$	$r = 0$	$r <= 2$	$r <= 1$	$r = 0$	n_{US}	n_j
Canada	7		0.24	8.64	40.87***	0.24	8.40	32.22***	12.85	16.01
									(2.51)	(2.97)
France	12	$\surd$	3.77	16.52	41.69*	3.77	12.75	25.17*	7.75	1.01
									(1.15)	(0.98)
Germany	7	$\surd$	4.53	16.04	42.28*	4.53	11.51	26.24**	4.26	-1.42
									(0.67)	(0.23)
Japan	7	$\surd$	4.92	16.91	34.79	4.92	11.99	17.88	—	—
									(—)	(—)
Sweden	8	$\surd$	7.28	18.24	63.74***	7.28	10.96	45.50***	16.57	9.43
									(2.49)	(1.68)
United Kingdom	3		0.01	11.09	41.75***	0.01	11.08	30.66***	5.05	4.59
									(1.21)	(1.17)

NOTE: The system variables are $(q_t, c_{t,US}, c_{t,j})$, where q_t is the real exchange rate and $c_{t,k}$ is the real consumption at time t for country $k = i, j$. The cointegrating vector has been normalised with respect to the real exchange rate. Figures in parentheses denote standard errors. The lag length of the Vector Autoregression (VAR) was determined on the basis of the Schwartz Information Criterion (SIC), allowing for a maximum length of 8 lags. However, in the case that the LM test indicated that the VEC residuals exhibited serial correlation up to order 8 the lag length was increased. ***, **, * denote significance at the 1%, 5%, and 10% significance level, respectively.

This is illustrated in Table 4.2 which reports correlation coefficients between the real exchange rate and consumptions for three different cases. The first case (ρ) corresponds to the Backus and Smith (1993) model where the relative risk aversion coefficients are assumed to be identical and a time trend is not included in the equilibrium value. In the second case (ρ_{ra}), the assumption of identical risk aversion coefficients is relaxed by using the estimates from Table 4.1. Finally, we also consider the effect of different risk aversion coefficients and a time trend (ρ_{tra}) for the cases that the latter is significant in the cointegration analysis.

Table 4.2: The Consumption Real Exchange Rate Anomaly

Correlation	Canada	France	Sweden	United Kingdom
ρ	-0.575	0.101	-0.553	-0.274
ρ_{ra}	0.393	0.045	-0.556	0.031
ρ_{tra}	—	0.255	0.149	—

NOTE: ρ denotes the correlation coefficient between real exchange rate and relative consumption, while ρ_{ra} is the correlation coefficient adjusted for the different levels of relative risk aversion and ρ_{tra} is the coefficient adjusted for both the different levels of relative risk aversion and a time trend.

As far as Germany is concerned, although there is evidence of cointegration the results are not in line with IRBC models since the coefficient of relative risk aversion is negative. In summary, these findings support mean reversion towards the time-varying equilibrium specified by IRBC models.

4.3.2 Linearity Testing

A complementary reason for the empirical regularities reported in the IRBC literature, such as the estimated values of correlation coefficients (Backus and Smith, 1993), the difference in volatility (Obstfeld and Rogoff, 2000), and the difficulty of finding cointegration (Kollmann, 1995), may be that the deviations process is governed by nonlinear dynamics. Next, we investigate whether the deviations series exhibit significant STAR nonlinearity of the type suggested by Kilian and Taylor (2003). Escribano and Jordá (1999) developed a linearity test that provides useful insights concerning the presence of STAR nonlinearity, and the specification of the transition variable.

In deriving an LM test for the null of linearity against STAR nonlinearity we adopt the typical STAR model (Teräsvirta, 1994; Escribano and Jordá, 1999; van Dijk et al., 2002)

$$z_t = \boldsymbol{\phi}' \boldsymbol{x}_t + \boldsymbol{\theta}' \boldsymbol{x}_t F(s_t, \gamma, c) + u_t, \tag{4.8}$$

where $\boldsymbol{x}_t = (1, z_{t-1}, \ldots, z_{t-p})'$, $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_p)$, $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_p)$, s_t is the transition variable, γ is the transition parameter and c is a constant. The transition function $F(\cdot)$ for the ESTAR model is defined by

$$F(s_t, \gamma, c) = \left[1 - \exp\left(-\gamma (s_t - c)^2\right)\right]. \tag{4.9}$$

In the case of a Logistic STAR (LSTAR) model

$$F(s_t, \gamma, c) = \left[1 + \exp\left(-\gamma (s_t - c)\right)\right]^{-1}. \tag{4.10}$$

Testing linearity in this framework is not straightforward due to the presence of unidentified-nuisance parameters (Davies, 1977). Luukkonen et al. (1988) overcome the identification problem by replacing $F(\cdot)$ with a Taylor series approximation. The resulting equation permits the use of LM tests which asymptotically posses the χ^2 distribution. Escribano and Jordá (1999) extended the work of Luukkonen et al. (1988) and Teräsvirta (1994) and proposed a new specification strategy to choose between ESTAR and LSTAR models based on the following equation[7]

$$z_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \boldsymbol{x}_t s_t + \boldsymbol{\delta}_2' \boldsymbol{x}_t s_t^2 + \boldsymbol{\delta}_3' \boldsymbol{x}_t s_t^3 + \boldsymbol{\delta}_4' \boldsymbol{x}_t s_t^4 + u_t. \tag{4.11}$$

- Estimate Equation (4.11) and obtain the p-value, p_1, for the null hypothesis

[7]This new procedure appears to be consistent and to generate much higher correct selection frequencies (see Paya and Peel, 2005).

of linearity, $H_0^1 : \delta_1' = \delta_2' = \delta_3' = \delta_4' = 0$.

- If linearity is rejected,
  - test the null $H_0^L : \delta_2' = \delta_4' = 0$ with an F-test and obtain the corresponding p-value, p_E.
  - test the null $H_0^L : \delta_1' = \delta_3' = 0$ with an F-test and obtain the corresponding p-value, p_L.
- If $p_E < p_L$ select ESTAR, otherwise select LSTAR.

The implementation of the above procedure requires the specification of the lag length p and the transition variable s_t. We follow previous studies and set $p = 2$ for all countries but Sweden and the United Kingdom, for which we set $p = 4$ so as to deal with residual autocorrelation. Kilian and Taylor (2003) argue that although most studies employ a single past deviation as the transition variable, it is more intuitive to allow the effects of persistent deviations to be cumulative. To this end, we consider $s_t = (\sum_{d=1}^{\bar{d}} z_{t-d}^2)^{1/2}$, where $\bar{d}$ denotes the lag with the minimum p-value for the null of linearity, H_0^1, and we allow a maximum of 8 lags.

An important issue when testing the presence of STAR nonlinearities is the presence of conditional heteroskedasticity in the model's residuals. For example, Lundbergh and Teräsvirta (1998) examine the linearity test of Teräsvirta (1994) and conclude that conditional heteroskedasticity may result in severe size distortions and that the robust version of Granger and Teräsvirta (1993) appears to have very low power.[8] Pavlidis et al. (2009b) show that the Escribano and Jordá (1999) test exhibits similar problems as the test of Teräsvirta (1994) and investigate the performance of possible alternatives to improve its properties (size and size-adjusted power). Their findings suggest that the use of the Heteroskedasticity Consistent Covariance Matrix Estimator (HCCME) of MacKinnon and White (1985) improves upon the size, but results in very low size-adjusted power. On

[8]See Lundbergh and Teräsvirta (1998) for the specification, estimation and evaluation of models with nonlinear behaviour in the mean (STAR) and in the conditional variance (STGARCH), the STAR-STGARCH model.

the other hand, the Fixed Design Wild Bootstrap appears to lead to a marked improvement both in terms of size and size-adjusted power.[9]

Table 4.3: Linearity Testing

Country j		p	$\bar{d}$	p_1	p_E	p_L
Canada						
	OLS	2	8	0.002	0.008	0.003
	HC	2	8	0.023	0.165	0.023
	WB	2	8	0.008	0.028	0.010
France						
	OLS	2	7	0.066	0.029	0.017
	HC	2	7	0.051	0.030	0.017
	WB	2	7	0.080	0.052	0.030
Germany						
	OLS	2	3	0.298	0.133	0.118
	HC	2	2	0.235	0.122	0.109
	WB	2	3	0.364	0.226	0.194
Sweden						
	OLS	4	1	0.033	0.015	0.036
	HC	4	6	0.308	0.692	0.753
	WB	4	5	0.226	0.142	0.194
United Kingdom						
	OLS	4	2	0.006	0.014	0.012
	HC	4	2	0.001	0.040	0.030
	WB	4	2	0.036	0.044	0.040

NOTE: The length of the autocorrelation is denoted by p, while $\bar{d}$ shows the number of lags included in the transition variable for which the p-value for the null of linearity, p_1, is the lowest. p_E and p_L are the p-values for the null hypotheses of LSTAR and ESTAR nonlinearity, respectively.

The results of the Escribano and Jordá procedure using the Least Squares Covariance Matrix (LS), the Heteroskedasticity Consistent Covariance Matrix of MacKinnon and White (1985) (HC), and the Fixed Design Wild Bootstrap (WB) are presented in Table 4.3. Overall, linearity is rejected for the majority of cases. For Canada, France, and the United Kingdom the results are qualitatively similar between the three versions of the LM test. The null hypothesis, H_0^1, is rejected at least at the 10% significance level and the same transition variable is selected for each country by all tests, $\bar{d} = 8, 7$ and 2. The latter finding supports the use of the model proposed by Kilian and Taylor instead of the ESTAR model usually

[9]The Fixed Design Wild Bootstrap is described in detail in Chapter 2.

adopted in the literature with $\bar{d} = 1$. In the case of Sweden, only the original version of the Escribano and Jordá procedure rejects the null of linearity at the 5% significance level, which implies that $\bar{d} = 1$. Germany is the only country for which the deviations from the equilibrium do not appear to follow a STAR process. Although the results suggest the selection of an LSTAR model rather than an ESTAR for all countries but Sweden the difference in the associated p-values is marginal. Given that there is no prior reason for an asymmetric adjustment, the remaining analysis focuses on ESTAR models.

4.3.3 Estimation of the ESTAR models and the Wild Bootstrap

We examine the performance of the two ESTAR models (4.4) and (4.5) (discussed in Section 4.2.1) in capturing the nonlinear dynamics of the deviations series, z_t. While the former model is used for all countries, the model proposed by Kilian and Taylor is only employed for Canada, France and the United Kingdom. This is due to the linearity test results, which indicate that the effect of the deviations are not cumulative, i.e. $\bar{d} < 2$, for Germany and Sweden. Furthermore, we cannot reject the restriction that z_t follows a unit root process at the equilibrium, $H_0 : \sum \phi_p = 1$. Table 4.3 shows the estimates of the restricted ESTAR models, the standard error of the regression, the corresponding t-statistic, the Ljung-Box Q-statistic for serial correlation in the residuals and the LM test statistic (ARCH) for conditional heteroskedasticity up to lags 1 and 4. The Q-statistic does not indicate the presence of serial correlation in the regression residuals. However, there is some evidence of conditional heteroskedasticity for Sweden and the United Kingdom.

In order to test the significance of the smoothness parameter, γ, in the presence of conditional heteroskedasticity or non normality we employ the Fixed Design Wild Bootstrap (see, e.g., Wu, 1986; Mammen, 1993; Davidson and Flachaire, 2001). The asymptotic validity of the Fixed Design Wild Bootstrap for stationary autoregressions with known finite lag order when the error term exhibits conditional heteroskedasticity of unknown form is established in Gonçalves and Kilian (2004). Their results cover as special cases the N-GARCH, t-GARCH and asymmetric GARCH models, as well as, stochastic volatility models. The procedure we follow is to impose the null H_0: $\gamma = 0$ and simulate 1,000 series for z_t, denoted by z_t^b according to

$$z_t^b = \hat{\mu} + \sum_{p=1}^{\bar{p}} \hat{\phi}_p(z_{t-p} - \hat{\mu}) + \epsilon_t^b. \tag{4.12}$$

The residuals ϵ_t^b are constructed by multiplying the residuals obtained by the ESTAR model, $\hat{\epsilon}_t$, by a random variable, η_t, that follows the Rademacher distribution

$$\eta_t = \begin{cases} -1 & \text{with probability } p = 0.5\,, \\ 1 & \text{with probability } (1-p), \end{cases}$$

The η_t are mutually independent drawings from a distribution independent of the original data. The distribution has the properties that $E(\eta_t) = 0$, $E(\eta_t^2) = 1$, $E(\eta_t^3) = 0$, and $E(\eta_t^4) = 1$. A consequence of these properties is that any heteroskedasticity or symmetric non-normality in the estimated residuals ($\hat{\epsilon}_t$) is preserved in the newly created residuals.[10]

This procedure provides an empirical distribution for $\hat{\gamma}$ and the associated standard errors. The idea in $1,000$ replications is to determine the appropriate t-values so we do not reject the null of $\hat{\gamma} = 0$. These critical values can then be used to determine whether the estimates of $\hat{\gamma}$ reject the null or not (see also Paya and Peel, 2006a). The Wild Bootstrap p-values under the null H_0: $\gamma = 0$, are also reported in Table 4.4.

[10]The Wild Bootstrap matches the moments of the observed error distribution around the estimated regression function at each design point ($\hat{z}^b$). Liu (1988) and Mammen (1993) show that the asymptotic distribution of the Wild Bootstrap statistics are the same as the statistics they try to mimic.

Table 4.4: Estimated ESTAR Models

Canada

(a) Typical ESTAR Model Parameterisation

$$\hat{z}_t = \underset{[1.913]}{-0.287} + (\underset{[11.105]}{1.002}(z_{t-1} + \underset{[1.913]}{0.287}) + (1 - \underset{[11.105]}{1.002})(z_{t-2} + \underset{[1.913]}{0.287}))\cdot$$
$$\cdot\exp(-\underset{\substack{[3.217]\\(0.010)}}{0.069}(z_{t-1} + \underset{[1.913]}{0.287})^2)$$

$s = 0.146$, $Q_1 = 0.784$ (0.376), $Q_4 = 3.803$ (0.433),

$\mathrm{ARCH}_1 = 0.005$ (0.946), $\mathrm{ARCH}_4 = 0.579$ (0.678)

(b) Kilian and Taylor Parameterisation

$$\hat{z}_t = \underset{[1.783]}{-0.145} + (\underset{[9.467]}{0.893}(z_{t-1} + \underset{[1.783]}{0.145}) + (1 - \underset{[9.467]}{0.893})(z_{t-2} + \underset{[1.783]}{0.145}))\cdot$$
$$\cdot\exp(-\underset{\substack{[2.966]\\(0.000)}}{0.032}\Sigma_{d=1}^{8}(z_{t-d} + \underset{[1.783]}{0.145})^2)$$

$s = 0.137$, $Q_1 = 0.011$ (0.917), $Q_4 = 2.739$ (0.602),

$\mathrm{ARCH}_1 = 1.829$ (0.179), $\mathrm{ARCH}_4 = 0.684$ (0.605)

France

(a) Typical ESTAR Model Parameterisation

$$\hat{z}_t = \underset{[0.900]}{0.040} + (\underset{[14.423]}{1.382}(z_{t-1} - \underset{[0.900]}{0.040}) + (1 - \underset{[14.423]}{1.382})(z_{t-2} - \underset{[0.900]}{0.040}))\cdot$$
$$\cdot\exp(-\underset{\substack{[3.153]\\(0.000)}}{0.643}(z_{t-1} - \underset{[0.900]}{0.040})^2)$$

$s = 0.074$, $Q_1 = 0.242$ (0.623), $Q_4 = 3.378$ (0.497),

$\mathrm{ARCH}_1 = 0.087\ (0.769)$, $\mathrm{ARCH}_4 = 0.294\ (0.881)$

(b) Kilian and Taylor Parameterisation

$$\hat{z}_t = \underset{[1.403]}{0.066} - (\underset{[12.945]}{1.310}(z_{t-1} - \underset{[1.403]}{0.066}) + (1 - \underset{[12.945]}{1.310})(z_{t-2} - \underset{[1.403]}{0.066}))\cdot$$
$$\cdot \exp(-\underset{\substack{[2.551]\\(0.003)}}{0.121}\,\Sigma_{d=1}^{7}(z_{t-d} - \underset{[1.403]}{0.066})^2)$$

$s = 0.077$, $Q_1 = 0.231\ (0.630)$, $Q_4 = 2.140\ (0.710)$,

$\mathrm{ARCH}_1 = 0.271\ (0.604)$, $\mathrm{ARCH}_4 = 0.602\ (0.662)$

Germany

(a) Typical ESTAR Model Parameterisation

$$\hat{z}_t = -\underset{[0.774]}{0.028} + (\underset{[11.797]}{1.190}(z_{t-1} + \underset{[0.774]}{0.028}) + (1 - \underset{[11.797]}{1.190})(z_{t-2} + \underset{[0.774]}{0.028}))\cdot$$
$$\cdot \exp(-\underset{\substack{[2.508]\\(0.000)}}{2.340}(z_{t-1} + \underset{[0.774]}{0.028})^2)$$

$s = 0.066$, $Q_1 = 0.293\ (0.588)$, $Q_4 = 2.138\ (0.710)$,

$\mathrm{ARCH}_1 = 1.582\ (0.211)$, $\mathrm{ARCH}_4 = 0.751\ (0.560)$

Sweden

(a) Typical ESTAR Model Parameterisation

$$\hat{z}_t = \underset{[0.152]}{0.010} + (\underset{[13.342]}{1.180}(z_{t-1} - \underset{[0.152]}{0.010}) - \underset{[0.709]}{0.093}(z_{t-2} - \underset{[0.152]}{0.010}) + \underset{[1.536]}{0.200}\cdot$$
$$\cdot(z_{t-3} - \underset{[0.152]}{0.010}) + (1 - \underset{[13.342]}{1.180} + \underset{[0.152]}{0.093} - \underset{[1.536]}{0.200})(z_{t-4} - \underset{[0.152]}{0.010}))\cdot$$
$$\cdot \exp(-\underset{\substack{[3.770]\\(0.000)}}{0.305}(z_{t-1} - \underset{[0.152]}{0.010})^2)$$

$s = 0.159$, $Q_1 = 0.303\ (0.582)$, $Q_4 = 0.540\ (0.969)$,

$\text{ARCH}_1 = 4.046\ (0.047)$, $\text{ARCH}_4 = 1.136\ (0.343)$

United Kingdom

(a) Typical ESTAR Model Parameterisation

$$\begin{aligned}\hat{z}_t = & -\underset{[1.151]}{0.051} + (\underset{[12.547]}{1.134}(z_{t-1} + \underset{[1.151]}{0.051}) + \underset{[0.259]}{0.034}(z_{t-2} + \underset{[1.151]}{0.051}) + \underset{[0.590]}{0.078}\cdot \\ & \cdot(z_{t-3} + \underset{[1.151]}{0.051}) + (1 - \underset{[12.547]}{1.134} - \underset{[0.259]}{0.034} - \underset{[0.590]}{0.078})(z_{t-4} + \underset{[1.151]}{0.051}))\cdot \\ & \cdot\exp(-\underset{\substack{[3.276]\\(0.000)}}{0.652}(z_{t-1} + \underset{[1.151]}{0.051})^2)\end{aligned}$$

$s = 0.075$, $Q_1 = 0.045\ (0.832)$, $Q_4 = 2.379\ (0.666)$,

$\text{ARCH}_1 = 5.212\ (0.024)$, $\text{ARCH}_4 = 2.163\ (0.078)$

(b) Kilian and Taylor Parameterisation

$$\begin{aligned}\hat{z}_t = & -\underset{[1.198]}{0.046} + (\underset{[12.233]}{1.093}(z_{t-1} + \underset{[1.198]}{0.046}) + \underset{[0.617]}{0.082}(z_{t-2} + \underset{[1.198]}{0.046}) + \underset{[0.725]}{0.095}\cdot \\ & \cdot(z_{t-3} + \underset{[1.198]}{0.046}) + (1 - \underset{[12.233]}{1.093} - \underset{[0.617]}{0.082} - \underset{[0.725]}{0.095})(z_{t-4} + \underset{[1.198]}{0.046}))\cdot \\ & \cdot\exp(-\underset{\substack{[3.684]\\(0.000)}}{0.387}\Sigma_{d=1}^{2}(z_{t-d} + \underset{[1.198]}{0.046})^2)\end{aligned}$$

$s = 0.074$, $Q_1 = 0.008\ (0.930)$, $Q_4 = 1.645\ (0.897)$,

$\text{ARCH}_1 = 4.936\ (0.028)$, $\text{ARCH}_4 = 1.801\ (0.133)$

NOTE: Figures in square brackets denote the ratio of the absolute value of the estimated coefficient to the estimated standard error of the coefficient estimate. The Wild Bootstrap $p-$values for the γ coefficient are reported in parentheses below the coefficient estimates. s is the standard error of the regression. Q_1 and Q_4 denote the Ljung-Box Q-statistic for serial correlation up to order 1 and 4, respectively. ARCH_1 and ARCH_4 denote the LM test statistic for conditional heteroskedasticity up to order 1 and 4, respectively.

The Wild Bootstrap p-values imply that the estimated transition parameters are in each case significant for all conventional levels, which supports the nonlinear nature of the deviation processes. Therefore, the difficulty of detecting cointegration in short samples may be attributed to large and persistent deviations generated by the ESTAR adjustment mechanism. Further, the high short term volatility of the real exchange rates compared to the volatility of the consumption series is, also, in accordance with the implications of the ESTAR model. We conduct a Monte Carlo experiment in the which illustrates the above points.

4.3.4 Generating the Puzzle

We are interested in examining the behaviour of the correlation coefficients between the real exchange rate and relative consumption and the properties of linear cointegration tests when the true DGP is nonlinear.

To this end, we calibrate nonlinear models by using parameter values similar to the estimated ones. For simplicity we assume that the two consumption series follow a driftless random walk

$$
\begin{aligned}
c_{i,t} &= c_{i,t-1} + u_{i,t}, \qquad u_{i,t} \sim N(0, 0.02), \\
c_{j,t} &= c_{j,t-1} + u_{j,t}, \qquad u_{j,t} \sim N(0, 0.02),
\end{aligned}
$$

while, the DGP for the real exchange rate is given by

$$
\begin{aligned}
q_t \;=\;& -0.07t + 6c_{i,t} - 9c_{j,t} + (1.2(q_{t-1} + 0.07(t-1) - 6c_{i,t-1} + 9c_{j,t-1}) - \\
& -0.1(q_{t-2} + 0.07(t-2) - 6c_{i,t-2} + 9c_{j,t-2}) + 0.2(q_{t-3} + 0.07(t-3) - \\
& -6c_{i,t-3} + 9c_{j,t-3}) + 0.1(q_{t-4} + 0.07(t-4) - 6c_{i,t-4} + 9c_{j,t-4})) \cdot \\
& \cdot \exp\left(-0.3(q_{t-1} + 0.07(t-1) - 6c_{i,t-1} + 9 * c_{j,t-1})^2\right) + \epsilon_t,
\end{aligned}
$$

where $\epsilon_t \sim N(0, 0.15)$. We set the sample size equal to 128 observations and generate 1,000 series for each variable. In turn, we obtain the correlation coefficients between the "fake" real exchange rate series and the "fake" relative consumption. The percentage of negative correlation coefficients is 46.4, implying that the likelihood of observing a small or negative correlation is large.

Further, we examine the power of the Johansen (1991) test to detect cointegration between the "fake" q_t, $c_{i,t}$ and $c_{j,t}$. The null hypothesis of no cointegration can be rejected in 43.1 percent of the cases when the nominal significance level is 10 percent. However, if we change the sample size to 70 observations, which is about the sample size used by Kollmann (1995), the power deteriorates to only 15.7 percent, indicating the importance of the sample length.

4.3.5 Generalized Impulse Response Functions

In this context, it is also of importance to investigate whether the estimated nonlinear models, as well as, the inclusion of the equilibrium determinants can explain the PPP puzzle regarding the slow rate at which shocks appear to damp out. Impulse response analysis addresses this issue by focusing on the effect of a shock on the behaviour of the deviation process. However, a number of studies have shown that impulse response analysis is considerably more complex for nonlinear models when compared to linear models (see Gallant et al., 1993; Koop et al., 1996; Potter, 2000; van Dijk et al., 2007). In particular, impulse responses produced by nonlinear models are history dependent, so they depend on initial conditions; they are dependent on the size and sign of the current shock; and they depend on the shocks that occur in future periods. Koop et al. (1996) propose a measure, the Generalized Impulse Response Function (GIRF), which deals with the complications entailed in impulse response analysis for nonlinear models. The GIRF is defined as the average difference between two realizations of the stochastic process, z_{t+h}, which start with identical histories up to time $t-1$, but only the first realization is hit by a shock of magnitude δ_t at period t.

$$\text{GIRF}(h, \delta_t, \omega_{t-1}) = E\left[z_{t+h}|\epsilon_t = \delta_t, \omega_{t-1}\right] - E\left[z_{t+h}|\omega_{t-1}\right], \tag{4.13}$$

where $h = 1, 2\ldots$ denotes horizon, $\epsilon_t = \delta_t$ is an arbitrary shock occurring at time t, and ω_{t-1} is the history set of z_t. Given that the GIRF$(h, \delta, \omega_{t-1})$ is a function of δ_t and ω_{t-1}, which are realizations of random variables, the GIRF$(h, \delta, \omega_{t-1})$ itself is a realization of a random variable. It follows that various conditional versions of the GIRF can be defined. For example, we can condition on the shock and treat the variables generating the history as random. Alternatively, we can consider a specific history and treat the GIRF as a random variable in terms of the shock. In general, we can condition on a subset of shocks and a subset of histories, depending on the specific application. In this work we choose to condition upon 'all past histories' so as to examine the time profile of the effects of shocks of different magnitudes on the future patterns of the series variable.

Due to the fact that analytic expressions for the conditional expectations involved in (4.13) are usually not available for $h > 1$, we use bootstrap integration methods (see Koop et al., 1996, for a detailed description) to overcome the issue of future shocks intrinsically incorporated in the model.[11] In particular, for each available history 200 repetitions are implemented to average out future shocks, where future shocks are drawn with replacement from the models residuals, and then the results across all histories are averaged. The maximum impulse response horizon is set to 48 quarters and we consider shocks of magnitude $\delta_t = \psi\hat{\sigma}_\epsilon$, where $\hat{\sigma}_\epsilon$ is the residual standard deviation and $\psi = 1, 3, 5$.

In order to measure the rate at which the final effect of an impulse, δ_t, is attained we compute the π-life or π-absorption time (see van Dijk et al., 2007)

$$N(\pi, \delta_t, \omega_{t-1}) = \sum_{m=0}^{\infty}\left(1 - \prod_{h=m}^{\infty} I(\pi, h, \delta_t, \omega_{t-1})\right), \tag{4.14}$$

[11] An analytical expression of the "impulse response function" for the deterministic skeleton of a restricted ESTAR model is provided by Venetis et al. (2007).

Table 4.5: Estimated π-lives of Shocks

		Shock Absorption		
Country	δ_t	0.25	0.50	0.80
	$1 \times \hat{\sigma}_\epsilon$	6 (4)	14 (10)	34 (22)
Canada	$3 \times \hat{\sigma}_\epsilon$	5 (4)	11 (8)	30 (18)
	$5 \times \hat{\sigma}_\epsilon$	3 (3)	9 (6)	25 (13)
	$1 \times \hat{\sigma}_\epsilon$	7 (7)	10 (9)	17 (15)
France	$3 \times \hat{\sigma}_\epsilon$	7 (6)	10 (9)	17 (15)
	$5 \times \hat{\sigma}_\epsilon$	5 (5)	8 (8)	15 (13)
	$1 \times \hat{\sigma}_\epsilon$	4 (-)	7 (-)	13 (-)
Germany	$3 \times \hat{\sigma}_\epsilon$	3 (-)	5 (-)	12 (-)
	$5 \times \hat{\sigma}_\epsilon$	2 (-)	3 (-)	9 (-)
	$1 \times \hat{\sigma}_\epsilon$	7 (-)	9 (-)	12 (-)
Sweden	$3 \times \hat{\sigma}_\epsilon$	6 (-)	8 (-)	11 (-)
	$5 \times \hat{\sigma}_\epsilon$	4 (-)	6 (-)	9 (-)
	$1 \times \hat{\sigma}_\epsilon$	7 (9)	9 (11)	11 (14)
United Kingdom	$3 \times \hat{\sigma}_\epsilon$	5 (7)	7 (9)	10 (13)
	$5 \times \hat{\sigma}_\epsilon$	4 (5)	5 (7)	8 (11)

NOTE: The table reports the absorption time for the typical ESTAR parameterisation and the Kilian and Taylor parameterisation. Figures in parentheses correspond to the latter. In the cases of Germany and Sweden only the typical ESTAR parameterisation is employed.

where $0 \leq \pi \leq 1$ and $I(\pi, h, \delta_t, \omega_{t-1})$ is the indicator function which takes the value of 1 if at least a fraction $1 - \pi$ of the difference between the initial and ultimate effects of δ_t has been absorbed after h periods and 0 otherwise.[12] The π-life corresponds to the minimum horizon beyond which the difference between the impulse responses at all longer horizons and the ultimate response is less than or equal to the fraction π of the difference between the initial impact and the ultimate response. Note that the above definition of π-life differs from the definition usually adopted in the literature, which is the shortest horizon at which at least a fraction $1 - \pi$ of the initial effect, δ_t, has been absorbed. This is an appealing feature since monotonicity is not granted. That is, $I(\pi, h, \delta_t, \omega_{t-1}) = 1$ does not necessarily imply $I(\pi, h + j, \delta_t, \omega_{t-1}) = 1, \forall j > 0$.

[12]The indicator function is defined as

$$I(\pi, h, \delta_t, \omega_{t-1}) = I\left[\left|\text{GIRF}(h, \delta_t, \omega_{t-1}) - \text{GIRF}^{\infty}(\delta_t, \omega_{t-1})\right| \leq \pi \left|\delta_t - \text{GIRF}^{\infty}(\delta_t, \omega_{t-1})\right|\right]$$

Table 4.5 displays the π-lives of shocks for the estimated ESTAR models of the deviation series (see Table 4.4). The results reported further illustrate the nonlinear nature of the real exchange rate with time-varying equilibrium models, with the absorption time decreasing with the size of the shock. Moreover, the reduction in the time needed to absorb fraction $(1 - \pi)$ of different size shocks depends on the proportion $(1 - \pi)$. In other words, if the shock increases from $1\hat{\sigma}_\epsilon$ to $5\hat{\sigma}_\epsilon$ the reduction in the time needed to absorb 25% of both shocks is not generally the same as the reduction in time needed to absorb 50% of the shocks. The half-lives corresponding to the smallest shocks range between 7 and 14 quarters, while for the largest shocks the half-lives range between 3 and 9 quarters. The absorption time also depends on the specific ESTAR formulation. For Canada and France the absorption time is much smaller when the Kilian and Taylor ESTAR model is adopted, but not for the United Kingdom. However, the results are qualitatively similar. Given that consensus estimates of linear models suggest a half-life between 3 and 5 years (see Rogoff, 1996), these results, in accordance with the results of other studies adopting a nonlinear framework, seem to go some way towards solving the PPP puzzle.

4.4 Conclusion

The present study adopts an IRBC framework, where the equilibrium real exchange rate is determined by consumption series. By focusing on the recent float, we find evidence in favour of a long-run relationship in line with the risk sharing condition implied by IRBC models with complete markets for most of the countries under examination. The results of linearity tests indicate that the deviations from the equilibrium, as estimated by the Johansen (1991) method, exhibit STAR nonlinearity. We fit ESTAR models and employ the Fixed Design Wild Bootstrap so as to draw inferences in the presence of conditional heteroskedasticity. The estimated models appear to parsimoniously fit the deviation processes. The nonlinear nature of the series provides an explanation for the empirical regularities noted in literature, as well as, the discouraging results reported for shorter spans of data. Finally, we address the PPP puzzle regarding the slow absorption rate of shocks by employing GIRFs. Our findings suggest that shocks to the deviations from the IRBC equilibrium have short half-lives.

CHAPTER 5

Real Exchange Rates and Time-Varying Trade Costs

The difficulty lies not so much in developing new ideas as in escaping from old ones.

John Maynard Keynes (1883 – 1946)

5.1 Introduction

Trade costs can exhibit significant economic magnitudes and can play an essential role in addressing several major puzzles in international economics (Obstfeld and Rogoff, 2000; Anderson and van Wincoop, 2004). In the Purchasing Power Parity (PPP) framework, equilibrium models of real exchange rate determination demonstrate how trade costs induce nonlinear but mean reverting adjustment toward PPP and, hence, provide a possible explanation for the well-documented persistence in the real exchange rate (Dumas, 1992; O'Connell and Wei, 2002; Taylor and Taylor, 2004). For example, O'Connell and Wei (2002) extend the iceberg model of trade to allow for fixed as well as proportional costs of arbitrage. As a consequence, the tendency of the real exchange rate to return to the equilibrium rate will become apparent only for misalignments which cover the level of transactions costs and imply arbitrage opportunities. Small misalignments, close to equilibrium and within the transactions band, will be left uncorrected so that the real exchange rate will exhibit near unit root behaviour.

In a number of empirical contributions trade costs are assumed constant and the implied type of nonlinear behaviour of the real exchange rate is modeled by the Exponential Smooth Transition Autoregressive (ESTAR) model (see, e.g., Michael et al., 1997; Kilian and Taylor, 2003; Taylor, Peel and Sarno, 2001). However, it can be argued that this assumption is too restrictive over long time periods.[1] In a recent study, inspired by the gravity literature, Jacks et al. (2008) present an aggregate micro-founded model which allows the construction of long span trade costs series. The authors illustrate that trade costs related to the exchange of goods across countries, far from been constant, have exhibited substantial and nonmonotonic changes from 1870 to 2000.[2] This finding has potentially important implications concerning the behaviour of the real exchange rate. Because trade costs vary in time so does the speed of mean reversion for a given PPP deviation (see, e.g., Dumas, 1992; Sercu et al., 1995). Intuitively, when trade costs increase (decrease) the trade costs band–in which no trade takes place– widens (narrows) and the real exchange rate process becomes more (less) persistent. Hence, the persistence of the real exchange rate does not only depend on the size of the deviation but also on the level of trade costs at each particular point in time. Neglecting significant changes in trade costs leads to underestimating/overestimating the degree of persistence and the time required for the process to absorb shocks at specific periods.

[1] Clemens and Williamson (2001) and Mohammed and Williamson (2004) among others illustrate that tariffs and global freight rates have fluctuated substantially in the last century. These studies focus on specific impediments of trade costs and, therefore, provide indirect evidence of time-varying trade costs. A survey on recent developments in the measurement of total trade costs and their components is provided by Anderson and van Wincoop (2004).

[2] Consequently, the effect of trade costs cannot be approximated by deterministic trends.

The contribution of this chapter is to report estimates and the properties of two smooth transition regression models of the real exchange rate which incorporate time-varying trade costs. The models are fitted to a long span of data (1830-2005) for the dollar-sterling real exchange rate and the trade costs index for the United Kingdom-United States country pair. Our choice is based on the fact that the relationship between trade frictions and the persistence of the real exchange rate should become apparent over long time periods in which large fluctuations of trade costs occur.

The rest of the chapter is structured as follows. In Section 5.2, we present the trade costs measure of Jacks et al. (2008). Section 5.3 outlines our nonlinear models of the real exchange rate. Section 4 deals with the description of the data and the empirical results. A summary and concluding comments are offered in the last section.

5.2 Trade Costs

> *"Trade costs, broadly defined, include all costs incurred in getting a good to a final user other than the marginal cost of producing the good itself"*
>
> Anderson and van Wincoop (2004, p. 691).

Obviously, trade costs break down into a vast number of components such as transportation costs (freight rates and time costs), policy barriers (tariffs and non-tariff barriers), informational costs and costs associated with the use of different currencies. The fact that several of these components are unobservable and data limitations pose serious problems in obtaining accurate estimates of the magnitude of total trade costs by direct atheoretical measures. The gravity literature circumvents this obstacle on the basis of theoretical models which enable measuring the degree of trade restrictiveness by extracting information from trade flows.

In this framework, Jacks et al. (2008) present a micro-founded measure of aggregate bilateral trade costs that captures trade frictions. The key idea in the derivation of their measure is that changes in trade barriers have an effect on both international and intranational trade. By establishing a relationship between countries' average international trade barriers and intranational trade, trade costs can be obtained directly from observable trade data without imposing a particular trade cost function (Novy, 2008).

Consider a world consisting of N countries and a continuum of differentiated goods. Anderson and van Wincoop (2003) derive the following gravity equation of international trade

$$x_{i,j} = \frac{y_i y_j}{y_w} \left(\frac{t_{i,j}}{\Pi_i P_j} \right)^{1-\sigma}, \tag{5.1}$$

where $x_{i,j}$ are nominal exports from country i to j. Income levels of country i, country j and world income are denoted by y_i, y_j and y_w, respectively. The elasticity of substitution, σ, is assumed to be constant and greater than unity. The cost of importing a good or, equivalently, the trade cost barrier (one plus the tariff equivalent) is $t_{i,j} \geq 1$. Finally, the price indices (or outward and inward multilateral resistance variables) Π_i and P_j for countries i and j represent the average trade restrictiveness of the countries. Novy (2008) uses Equation (5.1) to obtain a bidirectional gravity equation, which includes inward and outward multilateral resistance variables for both countries,

$$x_{i,j} x_{j,i} = \left(\frac{y_i y_j}{y_w} \right)^2 \left(\frac{t_{i,j} t_{j,i}}{\Pi_i P_j \Pi_j P_i} \right)^{1-\sigma}. \tag{5.2}$$

In turn, the author makes use of the fact that intranational trade, like international trade, depends on the magnitude of trade barriers, $x_{i,i} = ((y_i y_i)/y_w)(t_{i,i})/(\Pi_i P_i)^{1-\sigma}$, so as to control for multilateral resistance. Substituting into the bidirectional gravity equation yields

$$x_{i,j} x_{j,i} = x_{i,i} x_{j,j} \left(\frac{t_{i,j} t_{j,i}}{t_{i,i} t_{j,j}} \right)^{1-\sigma}. \tag{5.3}$$

The geometric average of the tariff equivalent can now be obtained by

$$\tau \equiv \left(\frac{t_{i,j}t_{j,i}}{t_{i,i}t_{j,j}}\right)^{\frac{1}{2}} - 1 = \left(\frac{x_{i,i}x_{j,j}}{x_{i,j}x_{j,i}}\right)^{\frac{1}{2(\sigma-1)}} - 1. \tag{5.4}$$

The above equation states that a drop in trade flows between countries with respect to trade flows within countries is associated with higher trade costs. Note that the micro-founded measure evaluates bilateral trade costs against the domestic trade cost benchmark. Further, it enables the construction of long-span trade costs series since its estimation only requires data for bilateral exports and intranational trade. The latter variable can be approximated by subtracting aggregate exports from a country's Gross Domestic Product (GDP) (Jacks et al., 2008).

5.3 Nonlinear Adjustment & Time-Varying Trade Costs

Let us define the log real exchange rate as $q_t = s_t - p_t + p_t^*$, where s_t is the logarithm of the spot exchange rate (the domestic price of foreign currency), p_t is the logarithm of the domestic price level and p_t^* the logarithm of the foreign price level.

5.3.1 The ESTAR Model

A widely employed nonlinear econometric model that can capture the behaviour of the real exchange rate in the presence of constant trade costs is the Exponential STAR (ESTAR) model advocated by Teräsvirta (1994). The appealing feature of the ESTAR model is that it allows transitions between a continuum of regimes to occur smoothly and symmetrically. In this setting, the speed of mean reversion is an increasing function of the size of the absolute deviation from equilibrium. This property is suggested by the analysis of Dumas (1992) and demonstrated by Berka (2005). In addition, Teräsvirta (1994) argues that if an aggregated process is observed, regime changes may be smooth rather than discrete as long as heterogeneous agents do not act simultaneously even if they individually make dichotomous decisions. All the above favour the use of ESTAR models over Threshold Autoregressive (TAR) models, in which changes of persistence occur abruptly.[3]

A STAR model for the process $\{q_t\}$ may be written as

$$q_t - \mu = \sum_{p=1}^{\bar{p}} \phi_p (q_{t-p} - \mu) G_j(\cdot) + \epsilon_t, \tag{5.5}$$

where μ is a constant representing the long run equilibrium, ϵ_t is a white noise process with mean 0 and variance σ_ϵ, and $G_j(\cdot)$ is the transition function. For a given AR structure, $\sum_{p=1}^{\bar{p}} \phi_p$, the transition function, $G_j(\cdot)$, specifies the degree of persistence of the real exchange rate at each point in time. In the presence of constant trade costs, the transition function for the ESTAR model is given by

$$G_1(q_{t-d}) = \exp\left(-\gamma^2 \left(q_{t-d} - \mu\right)^2\right) \tag{5.6}$$

where q_{t-d} is the transition variable and $\gamma > 0$ is the smoothness (or transition) parameter. The exponential transition function G_1 is particularly applicable because it implies symmetric adjustment for positive and negative deviations from the equilibrium. Furthermore, the speed of adjustment is increasing with the smoothness parameter γ and the absolute value of the past deviation from the equilibrium. For expositional reasons, we assume that $\sum_{p=1}^{\bar{p}} \phi_p = 1$ throughout this section. In this case, at the equilibrium $G_1(\cdot) = 1$ and the real exchange rate behaves as a unit root process, $q_t = \sum_{p=1}^{\bar{p}} \phi_p(q_{t-p} - \mu) + \epsilon_t$. Whilst, for nonzero deviations $G_1(\cdot) \in [0, 1)$ and the process becomes mean reverting. Finally, if $|q_{t-d} - \mu| \to \infty$ the function value approaches zero and the process is white noise, $q_t = \epsilon_t$. The speed of transition between regimes is specified by the smoothness parameter γ. If γ is equal to zero the real exchange rate behaves as a linear unit root process irrespectively of the regime. Whilst, if $\gamma \to \infty$ the process

[3]Note also that the incorporation of trade costs in TAR models is not straightforward.

becomes white noise. Intermediate values of γ imply smooth adjustment of the real exchange rate.

Let us consider two deviations from PPP which have the same size but occur at different time periods, $|q_{t_1-d} - \mu| = |q_{t_2-d} - \mu| \neq 0$ with $t_1 < t_2$. The fact that γ is constant in the typical ESTAR model implies that the real exchange rate will exhibit the same degree of persistence at time t_1 and t_2. Conditional on the assumption of constant trade costs this is an attractive property. However, if trade costs vary in time so will the speed of adjustment. An increase (decline) in trade costs, τ, during the two time periods, $\tau_{t_1-d} \neq \tau_{t_2-d}$, will induce higher (lower) persistence of the real exchange rate and, therefore, a decrease (increase) of the γ parameter. Hence, time varying trade costs can be incorporated into Equation (5.6) by allowing γ to change over time depending on τ_{t-d}. By assuming a linear relationship between the value of the smoothness parameter and trade costs, the transition function for the Time Varying Trade Costs ESTAR (TVTC-ESTAR) is given by

$$G_2(q_{t-d}, \tau_{t-d}) = \exp\left(-(\gamma - \gamma_\tau \tau_{t-d})^2 \left(q_{t-d} - \mu\right)^2\right), \tag{5.7}$$

where the coefficient, γ_τ, on trade costs is greater than zero and $\gamma \geq \gamma_\tau \tau_{t-d}$ $\forall$ t. The above equation allows both the degree of trade restrictiveness and the size of the deviation from the equilibrium to determine the speed of adjustment of the real exchange rate at a particular point in time (see Figure 5.1).

5.3.2 The QLSTAR Model

An alternative model to the ESTAR that captures the theoretical insights of the authors above and allows us to parsimoniously encompass the influence of fixed and proportional time-varying trade costs is the Quadratic Logistic Smooth Transition Autoregressive (QLSTAR) model of Jansen and Teräsvirta (1996). The transition function of the QLSTAR model is given by

$$G_3^\star(q_{t-d}) = 1 - \left(1 + \exp\left(-\gamma^2(q_{t-d} + c_1)(q_{t-d} + c_2)\right)\right)^{-1}, \tag{5.8}$$

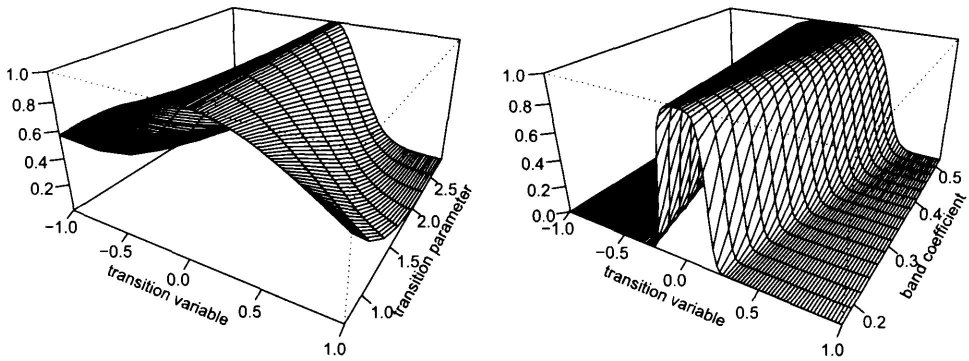

Figure 5.1: The exponential transition function (left) for $0.75 \leq \gamma - \gamma_\tau \tau_{t-d} \leq 3$, $q_{t-d} \in \{-1, \ldots, 1\}$, and $\mu = 0$. The quadratic logistic transition function (right) for $\gamma = 2.146$, $q_{t-d} \in \{-1, \ldots, 1\}$, $0.17 \leq c + c_\tau \tau_{t-d} \leq 0.52$, and $\mu = 0$.

where $c_1 = -\mu - c$ and $c_2 = -\mu + c$ with $c > 0$ are the band coefficients. The quadratic logistic transition function $G_3^\star(\cdot)$ is particularly applicable because it, as the exponential function, implies symmetric adjustment for positive and negative deviations from the equilibrium. Further, the QLSTAR model specified by Equation (5.8) can approximate ESTAR models but also nests three regime Threshold Autoregressive (TAR) models and linear AR models. In contrast to TAR and ESTAR models, the QLSTAR allows the type of adjustment (smooth or discrete) between regimes to be specified by the data and, at the same time, can approximate narrow and wide "bands of inaction". Hence, the model allows for both fixed and proportional costs. Overall, the model is particularly applicable when one is agnostic about the range of the "band of inaction" and the type of transition.

Suppose that regime changes occur abruptly rather than gradually (see Sercu et al., 1995), which favours the use of TAR over ESTAR models. If $\gamma \to \infty$ and $q_{t-d} < c_1$ or $q_{t-d} > c_2$ the transition function value equals zero and q_t becomes white noise. Whilst, inside the "band of inaction", $c_1 < q_{t-d} < c_2$, $G_3^\star(\cdot)$ equals one and q_t behaves as a unit root process. Note that an increase in trade costs will widen the "band of inaction" and, therefore, result in higher absolute values of the band coefficients, c_1 and c_2. At the other extreme, when $\gamma = 0$ the model becomes linear. For moderate values of γ, the QLSTAR model can approximate both ESTAR and TAR models. The speed of mean reversion increases with the absolute deviation from the equilibrium. If $|q_{t-d}-\mu| \to \infty$ the process approaches the white noise regime (outer regime). Whilst, in the inner regime, $q_{t-d} - \mu = 0$, the degree of persistence is given by the maximum value of the transition function $G_3^\star$

$$G_3^\star(\mu) = 1 - \left(1 + \exp\left(\gamma^2 c^2\right)\right)^{-1}, \tag{5.9}$$

which is determined by the transition parameter γ and the coefficient c. Consequently, changes in γ or c lead to different degrees of persistence at the equilibrium. Due to the fact that there is no *a priori* reason why changes in trade costs should alter the degree of persistence in the inner regime, we modify Equation (5.8) as follows

$$G_3(q_{t-d}) = 1 - \left(1 + \exp\left(-\frac{\gamma^2}{c^2}(q_{t-d} + c_1)(q_{t-d} + c_2)\right)\right)^{-1}. \tag{5.10}$$

The maximum value of $G_3(\cdot)$, which again occurs at the equilibrium rate, is

$$G_3(\mu) = 1 - \left(1 + \exp(\gamma^2)\right)^{-1}, \tag{5.11}$$

and is independent of the value of the band coefficient. The above modification enables the incorporation of time-varying trade costs in the QLSTAR model in a straightforward manner. The transition function for the Time-Varying Trade Costs QLSTAR (TVTC-QLSTAR) is given by

$$G_4(q_{t-d}, \tau_{t-d}) =$$
$$\left[1 - \left(1 + \exp\left(-\frac{\gamma^2}{(c + c_\tau \tau_{t-d})^2}(q_{t-d} + c_3)(q_{t-d} + c_4)\right)\right)^{-1}\right] \tag{5.12}$$

where $c_3 = -\mu - c - c_\tau \tau_{t-d}$ and $c_4 = -\mu + c + c_\tau \tau_{t-d}$ with $c_3 < c_4$ are the time-varying band coefficients, c is a positive constant, $c_\tau \geq 0$ is the coefficient on trade costs τ.[4] Controlling for γ, the speed of mean reversion decreases with

[4]We have scaled the trade costs index so as to have a minimum value of zero. Consequently, c

the absolute value of the band coefficients c_1 and c_2, and increases with the past deviation from the equilibrium rate (see Figure 5.1).[5] We examine the impact of trade costs on the speed of mean reversion of the real exchange rate in the next section.

5.4 Empirical Results

Our data set consists of annual observations for the dollar-sterling real exchange rate and the corresponding trade costs index from 1830 to 2005. For the construction of the real exchange rate we use the International Financial Statistics database to update the nominal exchange rate and the price indexes analysed in Lothian and Taylor (1996). International trade data are obtained by Mitchell (2008b,a) and GDP series for the United States and the United Kingdom are taken from Officer (2008) and Johnston and Williamson (2008), respectively.

Figure 5.2 shows the demeaned real exchange rate and the trade costs series. In line with Jacks et al. (2008), the latter exhibits significant fluctuations throughtout the period. Specifically, until the beginning of the $20th$ century trade costs were relatively low. Subsequently, the war and interwar periods were associated with a remarkable increase of bilateral trade costs with respect to intranational domestic costs. During this time interval the series displays two peaks, the first in 1935 following the Great Depression, and the second in 1946 at the end of the second World War and the establishment of the Bretton Woods system. A gradual decline has occurred since then.

After running a battery of linearity tests on the real exchange rate series, which indicate the presence of smooth transition nonlinearity, we examine whether trade costs are an important constituent of the nonlinear adjustment mechanism of the

reflects the lowest level of trade costs in time.

[5]Note that dividing the smoothness parameter γ^2 by $(c + c_\tau \tau_{t-d})^2$ also implies that changes in the persistence of the process become more abrupt as τ_{t-d} decreases. This behaviour is in line with the presence of both fixed and proportional costs which move together in time (O'Connell and Wei, 2002).

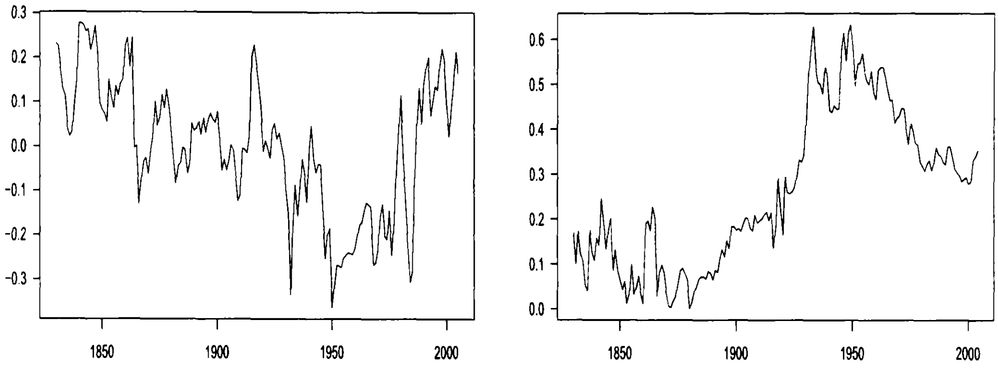

Figure 5.2: Time series plots of the demeaned dollar-sterling real exchange rate (left) and the United States-United Kingdom trade costs index (right).

real exchange rate.[6] The results for the nonlinear models with constant and time-varying trade costs are reported in Table 5.1.[7] Overall, all models provide a parsimonious fit to the real exchange rate. However, the incorporation of time-varying trade costs leads to a radically different adjustment process. The statistical significance of the coefficient γ_τ and the band coefficient c_τ of the TVTC-ESTAR and TVTC-QLSTAR models, respectively, indicates that movements in trade costs can help explain changes in the level of persistence of the real exchange rate.[8] An in-

[6]Specifically, we employ the testing procedures proposed by Teräsvirta (1994), Harvey and Leybourne (2007), and Kapetanios et al. (2003). The first two are general procedures for testing linearity against smooth transition nonlinearity. The main difference between them lies in the fact that the null critical values for the test of Teräsvirta (1994) are based on the assumption of an $I(0)$ process, whilst, the test of Harvey and Leybourne (2007) allows for both $I(0)$ and $I(1)$ processes. We find that the hypothesis of linearity can be rejected at the 5 and 10 percent significance levels, respectively. Finally, the test of Kapetanios et al. (2003) shows that the null hypothesis of a unit root in the real exchange rate against the alternative hypothesis of a globally stationary exponential smooth transition autoregressive process can be rejected at all conventional levels of significance. See also the results presented in Chapter 3.

[7]The models are fitted to the demeaned real exchange rate. The lag length of the autoregressive part and the variables which enter the transition function are specified on the basis of residual diagnostics and, subsequently, the statistical significance of the coefficients of the models. In the estimation procedure we impose the restriction $\phi_1 = 1$. This choice is based on the fact that the AR coefficient is not statistically different from unity in the estimated ESTAR models with constant and time-varying trade costs and in the TVTC-QLSTAR model. Further, the results for the unrestricted models are qualitatively the same. For the standard QLSTAR model imposing the restriction $\phi_1 = 1$ allows convergence of the nonlinear least squares algorithm. Note that this restriction does not necessarily imply a unit root behaviour of $\{q_t\}$ in the inner regime when QLSTAR models are applied since the maximum value of the transition function may differ from unity.

[8]Paya and Peel (2006a) emphasise that the high degree of persistence of both the dependent and explanatory variables (such as the trade costs series) that enter the transition function may give rise to a spurious regression problem. To this end, we report the bootstrap p-values for the coefficients on trade costs. The null Data Generating Process (DGP) in the simulation experiment is given by the fitted ESTAR and QLSTAR models.

crease in trade costs widens the "band of inaction" and reduces the speed of mean reversion for a given PPP deviation.

Table 5.1: Estimated Nonlinear Real Exchange Rate Models

Panel A, ESTAR

$$\hat{q}_t + \underset{(0.690)}{0.016} = (q_{t-1} + \underset{(0.690)}{0.016})\exp(-\underset{(7.102)}{1.505}^2(q_{t-1} + \underset{(0.690)}{0.016})^2).$$

$s = 0.064$; $Q_1 = 0.140$ [0.062]; $Q_5 = -0.127$ [0.227]; $\text{ARCH}_1 = 0.557$ [0.456]; $\text{ARCH}_5 = 0.802$ [0.550].

Panel B, TVTC-ESTAR

$$\hat{q}_t - \underset{(3.262)}{0.066} = (q_{t-1} - \underset{(3.262)}{0.066})\exp(-(\underset{(5.130)}{3.552} - \underset{\substack{(3.145)\\ [0.037]}}{5.324}\,\tau_{t-2})^2(q_{t-2} - \underset{(3.262)}{0.066})^2).$$

$s = 0.063$; $Q_1 = 0.035$ [0.642]; $Q_5 = -0.161$ [0.374]; $\text{ARCH}_1 = 1.538$ [0.217]; $\text{ARCH}_5 = 0.538$ [0.747].

Panel C, QLSTAR

$$\hat{q}_t + \underset{(0.656)}{0.014} = (q_{t-1} + \underset{(0.656)}{0.014})\Big[1 - \Big(1 + \exp(-\underset{(6.700)}{1.829}^2/\underset{(5.853)}{0.402}^2(q_{t-1} - 0.387)$$

$$\times(q_{t-1} + 0.416))\Big)^{-1}\Big].$$

$s = 0.064$; $Q_1 = 0.141$ [0.061]; $Q_5 = -0.126$ [0.219]; $\text{ARCH}_1 = 0.535$ [0.465]; $\text{ARCH}_5 = 0.786$ [0.561].

Panel D, TVTC-QLSTAR

$$\hat{q}_t - \underset{(4.064)}{0.059} = (q_{t-1} - \underset{(4.064)}{0.059})\Big[1 - \Big(1 + \exp(-\underset{(7.811)}{2.146}^2/(\underset{(6.929)}{0.172} + \underset{\substack{(4.488)\\ [0.008]}}{0.587}\,\tau_{t-2})^2$$

$$\times(q_{t-2} - 0.231 - \underset{\substack{(4.488)\\ [0.008]}}{0.587}\,\tau_{t-2})(q_{t-2} + 0.1128 + \underset{\substack{(4.488)\\ [0.008]}}{0.587}\,\tau_{t-2}))\Big)^{-1}\Big].$$

$s = 0.063$; $Q_1 = 0.020$ [0.787]; $Q_5 = -0.154$ [0.426]; $\text{ARCH}_1 = 0.667$ [0.411]; $\text{ARCH}_5 = 0.344$ [0.886].

Notes: Figures in parentheses and square brackets denote absolute t-statistics and p-values, respectively. The p-values for the coefficients on trade costs $\hat{\gamma}_\tau$ and $\hat{c}_\tau$ are obtained through a simulation exercise, where the bootstrap DGPs are the fitted ESTAR and QLSTAR models, respectively. For illustration purposes, we report the summation of the long run equilibrium estimate and the constant part of the band coefficients $\hat{\mu} \pm \hat{c}$. s is the standard error of the regression. Q_1 and Q_5 denote the Ljung-Box Q-statistic for serial correlation up to order 1 and 5, respectively. ARCH_1 and ARCH_5 denote the LM test statistic for conditional heteroskedasticity up to order 1 and 5, respectively.

Figure 5.3 displays the transition functions of the time-varying trade costs models for three representative time periods, namely 1900,1950 and 2000, which correspond to relatively low, large and moderate levels of trade costs, respectively. At those time periods, for the TVTC-ESTAR model, a PPP deviation of 0.4, which is roughly the maximum realized deviation, would suggest that the real exchange rate behaves similar to an AR process with coefficient around 0.3, a near unit root and an AR process with coefficient around 0.5. For the TVTC-QLSTAR model, the same PPP deviation would suggest that the real exchange rate behaves similar to a white noise, a near unit root and an AR process with coefficient around 0.2. On the other hand, according to the estimated ESTAR and QLSTAR models with constant trade costs the real exchange rate would behave as an AR process with coefficient of about 0.7 and 0.5, respectively, at all points in time. It appears that the inability of ESTAR models to approximate a wide "band of inaction" results in finding substantially higher persistence for large deviations than that implied by the QLSTAR.

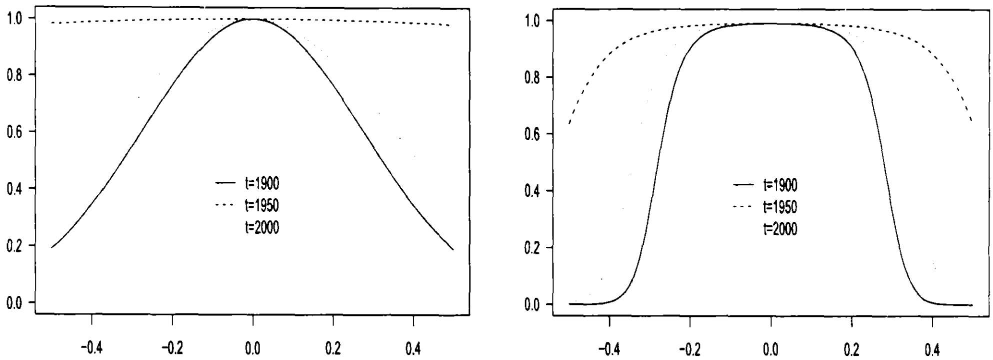

Figure 5.3: The exponential (left) and quadratic logistic (right) functions corresponding to 1900, 1950 and 2000 trade costs levels.

Clearly, the assumption of constant trade costs can result in severe overestimation / underestimation of persistence. The difference between the degrees of persistence (as measured by the value of the transition function of the corresponding model) estimated by the time-varying and constant trade costs models are illustrated in Figure 5.4. Starting with the ESTAR model, overestimation due to the the exclusion of time-varying trade costs occurs with almost the same likelihood as underestimation (55 percent versus 45 percent of the times). On the contrary, the QLSTAR model with constant trade costs appears to underestimate the degree of persistence with respect to the TVTC-QLSTAR in most periods (85 percent of the cases). Overestimation occurs on rare occasions (15 percent of the time) which are usually associated with substantial differences in the speed of mean reversion.[9]

A natural question that arises in the nonlinear framework is how fast does the process adjust to deviations from the equilibrium under different trade costs levels. In order to examine the time profile of the impact of a shock on the future behaviour of the series we adopt the Generalised Impulse Response Func-

[9]We note that the mean underestimation–the mean of the positive differences between the values of the transition function of the TVTC-ESTAR and the ESTAR– is 0.04 and the maximum value 0.24. While the mean overestimation–the mean of the negative differences between the values of the transition function of the TVTC-ESTAR and the ESTAR– is -0.07 and the minimum value is equal to -0.35. For the QLSTAR models, the mean underestimation is 0.04 and the maximum value 0.28. While the mean overestimation is -0.1 and the minimum value -0.48.

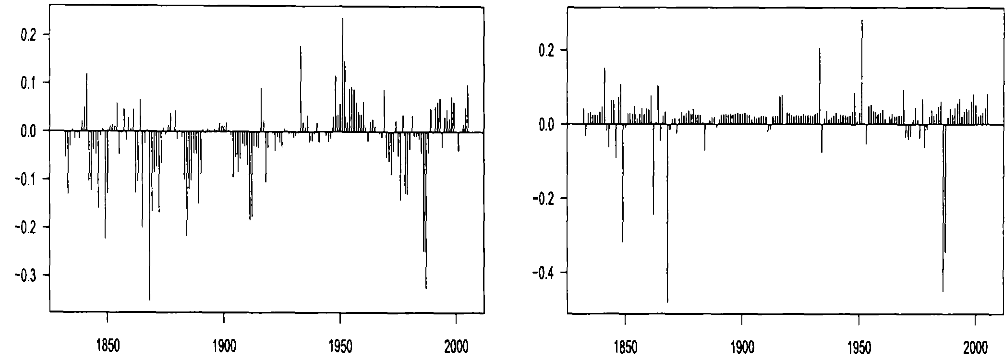

Figure 5.4: Differences in the degree of persistence between the TVTC-ESTAR and ESTAR models (left) and the degree of persistence between the TVTC-QLSTAR and QLSTAR models (right).

tions (GIRF) proposed by Koop et al. (1996).[10] The GIRF is defined as the average difference between two realizations of the stochastic process, q_{t+h}, which start with identical histories up to time $t-1$, but only the first realization is hit by a shock of magnitude δ_t at period t.

$$\text{GIRF}(h, \delta_t, \omega_{t-1}) = E\left[q_{t+h}|\epsilon_t = \delta_t, \omega_{t-1}\right] - E\left[q_{t+h}|\omega_{t-1}\right], \tag{5.13}$$

where $h = 1, 2 \ldots$ denotes horizon, $\epsilon_t = \delta_t$ is an arbitrary shock occurring at time t, and ω_{t-1} is the history set of q_t. Given that the GIRF$(h, \delta, \omega_{t-1})$ is a function of δ_t and ω_{t-1}, which are realizations of random variables, the GIRF$(h, \delta, \omega_{t-1})$ itself is a realization of a random variable. It follows that various conditional versions of the GIRF can be defined. In this work we set $\omega_{t-1} = \mu$, so that the process is initially at its equilibrium value, and we consider shocks of magnitude δ equal to the maximum absolute PPP deviation and half the maximum PPP deviation. Due to the fact that analytic expressions for the conditional expectations involved in (5.13) are usually not available for $h > 1$, we use bootstrap integration methods (see Koop et al., 1996, for a detailed description) to overcome the issue of future shocks intrinsically incorporated in the model. In particular, 1000 repetitions are implemented to average out future shocks, where future shocks are drawn with

[10] A more detailed description of the GIRF is provided in Chapter 4.

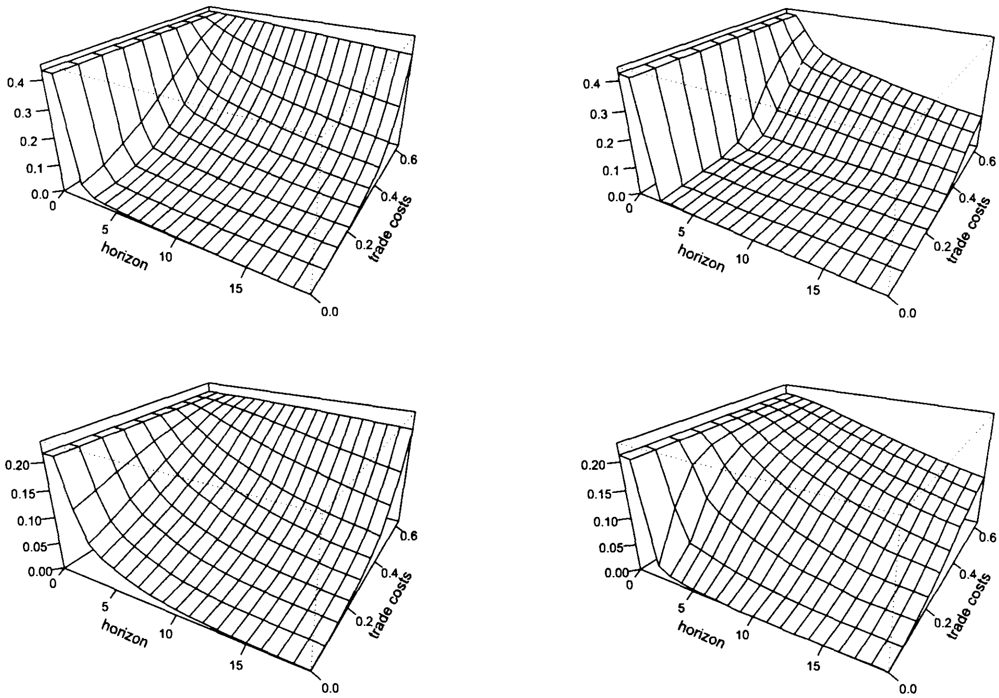

Figure 5.5: GIRFs for the TVTC-ESTAR (left) and TVTC-QLSTAR (right) models. Top (bottom) graphs correspond to shocks equal to the maximum absolute PPP deviation (half the maximum absolute PPP deviation).

replacement from the models residuals, and then the results are averaged.

Figure 5.5 illustrates the GIRFs for all levels of trade costs and for a maximum impulse response horizon of 20 years. Overall, low levels of trade costs are associated with fast shock absorption for all cases. The absorption time increases with the level of trade costs. For large shocks (maximum PPP deviation), the increase for the TVTC-ESTAR is substantially greater than for the TVTC-QLSTAR model and becomes apparent at a much lower level of trade costs. On the other hand, for moderate shocks (half the maximum PPP deviation), the absorption time for the TVTC-QLSTAR model initially grows faster as the degree of trade restrictiveness increases. However, this situation is reversed for high levels of trade costs. Generally, when the level of trade costs is high shocks fade out extremely slowly for the TVTC-ESTAR model. Put it differently, the transition parameter in the TVTC-ESTAR model approaches zero (infinite band width) falsely suggesting that the real exchange rate series is a unit root process.

To further illustrate this point as well as to make comparisons with the standard STAR models, we compute the half-lives corresponding to the maximum PPP deviation.[11] The results are presented in Table 5.2. Starting with the standard ESTAR and QLSTAR models, the real exchange rate process would absorb half of the shock in four years. Turning to the time-varying trade costs models, we consider three scenarios. Again, we set trade costs equal to their 1900, 1950 and 2000 levels. In the former and latter cases, both the TVTC-ESTAR and TVTC-QLSTAR models suggest that the time required for the process to absorb half of the maximum PPP deviation is only two years, which is half of that corresponding to constant trade costs. Obviously, large deviations of the real exchange rate appear to mean revert much faster (than that implied by the ESTAR and QLSTAR models) during the beginning of the $20th$ century and the recent floating period. On the contrary, the high level of trade costs around the middle of the $20th$ century leads to an increase in the half-life of the shock with respect to the constant trade costs benchmark. In particular, the TVTC-QLSTAR and TVTC-ESTAR models imply a half-life of 5 and 20 years, respectively. As above, the large discrepancy between the results of the two models can be attributed to the inability of the ESTAR model to capture the effect of wide "bands of inaction".[12]

Table 5.2: Half-Lives of the Nonlinear Real Exchange Rate Models

Trade Costs Level	ESTAR	QLSTAR	TVTC-ESTAR	TVTC-QLSTAR
1900	4	4	2	2
1950	4	4	12	5
2000	4	4	2	2

Notes: The size of the shock is set equal to the maximum PPP deviation. Half-lives are measured in years.

[11]The half-life is defined as to the minimum horizon beyond which the difference between the impulse responses at all longer horizons and the ultimate response is less than or equal to half of the difference between the initial impact and the ultimate response (van Dijk et al., 2007).

[12]We note that when trade costs reach a maximum, which occurs in 1946, the corresponding half-lives are 12 and 57 years for the TVTC-QLSTAR and TVTC-ESTAR models, respectively.

In order to examine which model is superior in terms of capturing the effect of time-varying trade costs, we conduct two bootstrap experiments. For each experiment, we employ either the estimated TVTC-QLSTAR or the TVTC-ESTAR model (reported in Table 5.1), the original trade costs series and the corresponding estimated residuals so as to generate B artificial samples of size 176.[13] In turn, we fit the alternative model to each artificial sample and compute the t-statistic for the coefficient on trade costs, $\tilde{t}_b$. This provides the empirical distributions for the t-statistics for $\hat{\gamma}_\tau$ and $\hat{c}_\tau$ under the null that the true DGP is given by the alternative model. The probability of obtaining a t-statistic as extreme as the original is

$$p_b = \frac{1}{B}\sum_{b=1}^{B} I(\tilde{t} \leq \tilde{t}_b),$$

where $I(A)$ is the indicator function, which takes the value of 1 if event A occurs and 0 otherwise, and $\tilde{t}$ is the original t-statistic. When the DGP is the TVTC-ESTAR model, the probability of the t-statistic for $\hat{c}_\tau$ exceeding 4.488 is only 13.8 percent. Whilst, when the DGP is given by the fitted TVTC-QLSTAR, there is a 52.1 percent probability that the value of the t-statistic for $\hat{\gamma}_\tau$ is greater than 3.145. Hence, it is very likely to obtain a t-statistic for the coefficient on trade costs in the TVTC-ESTAR model as extreme as the original when the DGP is given by the estimated TVTC-QLSTAR model. However, the opposite is not true.

5.5 Conclusion

In empirical work on the dynamic behaviour of the real exchange rates trade costs have typically been assumed constant. Essentially, arbitrage will commence, *ceteris paribus*, when it is profitable and PPP deviations are outside the transactions band. Motivated by the recent gravity literature we construct a long-span trade costs index. Further, we develop and estimate two nonlinear models for

[13]We set the number of generated samples B equal to 1000 and initialise the bootstrap DGP by using the first observations of the original real exchange rate series.

the real exchange rate which incorporate time-varying trade costs. Our empirical approach is supported by a battery of statistical tests and simulation methods. Our results provide strong evidence in favour of a time-varying "band of inaction", which widens with the level of trade costs. The persistence of the real exchange rate is found to depend on both the magnitude of trade frictions and the size of the deviation from PPP. For instance, a given shock to the real exchange rate would be absorbed at significantly different speeds in 1950 and 2000 due to the existence of different trade costs levels. Although trade costs appear to have declined substantially since the second World War, their magnitude is still significant. Consequently, our empirical results are also consistent with the documented high persistence of real exchange rates in the post-Bretton Woods era.

CHAPTER 6

Concluding Remarks

This thesis deals with the parsimonious modelling and forecasting of the real exchange rate using nonlinear econometric methods. In this context, the main research topics examined are: (i) robust linearity and unit root testing under the alternative of smooth transition nonlinearity, (ii) nonlinear real exchange rate model specification and forecasting, (iii) modelling the deviations of the real exchange rate from its fundamental value, and (iv) extending existing nonlinear real exchange rate models to accommodate for time-varying trade costs.

The first topic is addressed in Chapters 2 and 3. Chapter 2 deals with the specification stage of nonlinear models in the presence of conditional heteroskedasticity of unknown form. In particular, it investigates the impact of conditional heteroskedasticity on the performance of a conventional linearity test as well as several heteroskedasticity-robust versions. The key finding is that conventional tests can frequently result in the detection of spurious nonlinearity. The degree of oversizing depends on the type of time-varying volatility and tends to increase with the sample size. Consequently, spurious inference is most likely in high frequency data, such as daily and intra-daily financial time series. Conversely, when the true Data Generating Process is nonlinear in mean and the error is conditionally heteroskedastic, the tests may have very low size-adjusted power and can frequently lead to the selection of misspecified models. In most cases, the above deficiencies also hold for tests based on Heteroskedasticity Consistent Covariance Matrix Estimators. Overall, the Fixed Design Wild Bootstrap appears to be the most reliable method in terms of size, power and choosing the correct model specification. The importance of robust inference is highlighted through an empirical application to returns on major stock market indices and exchange rates, the future basis of the FTSE 100 and the dollar-sterling real exchange rate.

The following chapter extends the analysis to the nonlinear modelling and forecasting of the dollar-sterling real exchange rate using long spans of data. The contribution to the literature is threefold. First, we provide significant evidence of smooth transition dynamics in the series by employing a battery of recently developed in-sample statistical tests and bootstrap techniques. Second, we investigate through Monte Carlo simulations the small sample properties of several evaluation measures for comparing recursive forecasts when one of the competing models is nonlinear. Our results indicate that all tests exhibit low power in detecting the superiority of smooth transition over linear autoregressive models. Finally, notwithstanding the above, the nonlinear real exchange rate model outperforms both the random walk and the linear autoregressive model in forecasting the behaviour of the series during the post-Bretton Woods era. Consequently, researchers and practitioners can obtain forecasting gains regarding the behaviour of the long-span real exchange rate series by employing nonlinear models.

Chapter 4 adopts a more general framework, where the equilibrium real exchange rate is allowed to depend on the fundamentals implied by International Real Business Cycle models with complete asset markets. By focusing on the post-Bretton Woods era, we find that in several cases there is a long-run relationship between real exchange rates and consumption series in line with international risk sharing. Further, linearity tests indicate that the majority of the deviation processes exhibit significant smooth transition nonlinearity. Exponential Smooth Transition Autoregressive models parsimoniously capture the nonlinear adjustment mechanism. These findings provide an explanation for the empirical regularities noted in the literature on the relation between the real exchange rate and consumption, such as the "Backus and Smith (1993) puzzle". This point is illustrated further by generating the puzzle through Monte Carlo simulations. Finally, the results for Generalised Impulse Response functions show that shock absorption is significantly faster than suggested in the Purchasing Power Parity puzzle.

Chapter 5 takes a different approach from previous work on Purchasing Power Parity by explicitly accounting for time-varying trade costs. The motivation behind this approach is based on recent advances in the gravity literature which allow the construction of long-span trade costs indices. Our contribution is the development and estimation of two nonlinear models for the dollar-sterling real exchange rate which incorporate trade costs. The key finding is that both the magnitude of trade frictions and the size of the deviation from Purchasing Power Parity have a significant effect on the persistence of the real exchange rate. As a consequence, changes in trade costs imply that a given shock to the real exchange rate would be absorbed at substantially different speeds at different time periods. We provide evidence that the period after the Second World War was characterised by a substantial decline in the degree of trade restrictiveness. However, the magnitude of trade costs is still significant. Consequently, our empirical results are consistent with the documented high persistence of real exchange rates in the post-Bretton Woods era.

Bibliography

Anderson, J. E. and van Wincoop, E. (2003), "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93, 170–192.

— (2004), "Trade Costs," *Journal of Economic Literature*, 42, 691–751.

Apte, P., Sercu, P., and Uppal, R. (2004), "The Exchange Rate and Purchasing Power Parity: Extending the Theory and Tests," *Journal of International Money and Finance*, 23, 553–571.

Backus, D. K. and Smith, G. W. (1993), "Consumption and Real Exchange Rates in Dynamic Economies With Non-Traded Goods," *Journal of International Economics*, 35, 297–316.

Baillie, R. T. and Bollerslev, T. (1991), "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates," *Review of Economic Studies*, 58, 565–85.

— (2002), "The Message in Daily Exchange Rates: A Conditional-Variance Tale," *Journal of Business and Economic Statistics*, 20, 60–68.

Balassa, B. (1964), "The Purchasing Power Parity Doctrine: A Reappraisal," *Journal of Political Economy*, 72, 584–596.

Barro, R. (2005), "Rare Events and the Equity Premium," Working Paper 11310, National Bureau of Economic Research.

Benigno, G. and Thoenissen, C. (2008), "Consumption and Real Exchange Rates with Incomplete Markets and Non-traded Goods," *Journal of International Money and Finance*, 27, 926–948.

Bera, A. K. and Higgins, M. L. (1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence," *Journal of Business and Economic Statistics*, 15, 43–50.

Beran, R. (1988), "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687–697.

Berka, M. (2002), "General Equilibrium Model of Arbitrage Trade and Real Exchange Rate Persistence," Mimeo, University of British Columbia.

— (2005), "General Equilibrium Model of Arbitrage Trade and Real Exchange Rate Persistence," MPRA Paper 234, University Library of Munich, Germany.

Black, F. (1976), "Studies of Stock Price Volatility Changes," Proceedings of the American Statistical Association, Business and Economics Statistics Section, 177–181.

Blake, A. P. and Kapetanios, G. (2007), "Testing for ARCH in the Presence of Nonlinearity of Unknown Form in the Conditional Mean," *Journal of Econometrics*, 127, 472–488.

Boero, G. and Marrocu, E. (2004), "The performance of SETAR models: a regime conditional evaluation of point, interval and density forecasts," *International Journal of Forecasting*, 20, 305–320.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1993), "ARCH Models," University of California at San Diego, Economics Working Paper Series 93–49, Department of Economics, UC San Diego.

Campa, J. M. and Goldberg, L. S. (2002), "Exchange Rate Pass-Through into Import Prices: A Macro or Micro Phenomenon?" NBER Working Papers 8934, National Bureau of Economic Research, Inc.

Canzoneri, M., Cumby, R., and Diba, B. (1996), "Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence for a Panel of OECD Countries," *Journal of International Economics*, 47, 245–266.

Chan, F. and McAleer, M. (2002), "Maximum Likelihood Estimation of STAR and STAR-GARCH Models: Theory and Monte Carlo Evidence," *Journal of Applied Econometrics*, 17, 509–534.

Chari, V., Kehoe, P., and McGrattan, E. (2002), "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?" *Review of Economic Studies*, 69, 533–63.

Chinn, M. and Johnston, L. (1996), "Real Exchange Levels, Productivity and Demand Shocks: Evidence From a Panel of 14 Countries," NBER Working Paper No. 5709. Cambridge: MA.

Clark, T. E. (1999), "Finite-sample properties of tests for equal forecast accuracy," *Journal of Forecasting*, 18, 489–450.

Clark, T. E. and McCracken, M. W. (2001), "Tests of equal forecast accuracy and encompassing for nested models," *Journal of Econometrics*, 105, 85–110.

— (2005a), "Evaluating Direct Multistep Forecasts," *Econometric Reviews*, 24, 369–404.

— (2005b), "The power of tests of predictive ability in the presence of structural breaks," *Journal of Econometrics*, 124, 1–31.

Clemens, M. A. and Williamson, J. G. (2001), "A Tariff-Growth Paradox? Protection's Impact the World Around 1875-1997," Working Paper 8459, National Bureau of Economic Research.

Clements, M. P. (2005), *Evaluating Econometric Forecasts of Economic and Financial Variables*, New York: Palgrave Macmillan.

Clements, M. P. and Galvão, A. B. (2004), "A comparison of tests of nonlinear cointegration with application to the predictability of US interest rates using the term structure," *International Journal of Forecasting*, 20, 219–236.

Clements, M. P. and Smith, J. (1997), "The performance of alternative forecasting methods for SETAR models," *International Journal of Forecasting*, 13, 463–475.

— (1999), "A Monte Carlo Study of the Forecasting Performance of Empirical SETAR Models," *Journal of Applied Econometrics*, 14, 123–41.

Cook, S. (2006), "GARCH, heteroscedasticity-consistent covariance matrix estimation and (non)linear unit root testing," *Applied Financial Economics Letters*, 2, 217–222.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B., and Pictet, O. V. (2001), *An Introduction to High Frequency Finance*, San Diego: Academic Press.

Davidson, R. and Flachaire, E. (2001), "The Wild Bootstrap, Tamed at Last," Queen's Institute for Economic Research Working Paper No. 1000.

Davidson, R. and MacKinnon, J. G. (1985), "Heteroskedasticity–Robust Tests in Regression Directions," *Annales de l'INSEE*, 59/60, 183–218.

— (1999), "The Size Distortion Of Bootstrap Tests," *Econometric Theory*, 15, 361–376.

Davies, R. B. (1977), "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative," *Biometrika*, 64, 179–190.

— (1987), "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative," *Biometrika*, 74, 33–43.

De Grauwe, P., Dewachter, H., and Embrechts, M. (1993), *Exchange Rate Theory Chaotic Models of Foreign Exchange Markets*, Blackwell.

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

Deschamps, P. J. (2008), "Comparing Smooth Transition and Markov Switching Autoregressive Models of US Unemployment," *Journal of Applied Econometrics*, 23, 435–462.

Diebold, F. X. and Kilian, L. (2000), "Unit-Root Tests Are Useful for Selecting Forecasting Models," *Journal of Business & Economic Statistics*, 18, 265–73.

Diebold, F. X. and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–63.

Dumas, B. (1992), "Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World," *Review of Financial Studies*, 5, 153–80.

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

Efron, B. (1982), "The Jackknife, the Bootstrap, and Other Resampling Plans," monograph 38, CBMS-NSF, SIAM.

Eicker, F. (1963), "Asymptotic Normality of the Least Squares and Consistency of the Least Squares Estimators," *Annals of Mathematical Statistics*, 34, 447–456.

Eitrheim, Ø. and Teräsvirta, T. (1996), "Testing the Adequacy of Smooth Transition Autoregressive Models," *Journal of Econometrics*, 74, 59–75.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–36.

Engel, C. and Kim, C. (1999), "The Long-Run U.S./U.K. Real Exchange Rate," *Journal of Money Credit and Banking*, 31, 335–356.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

— (1990), "Stock Volatility and the Crash of '87: Discussion," *Review of Financial Studies*, 3, 103–06.

— (2001), "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics," *Journal of Economic Perspectives*, 15, 157–168.

Engle, R. F. and Granger, C. W. J. (1987), "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–76.

Engle, R. F. and Ng, V. (1993), "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–78.

Escribano, A. and Jordá, O. (1999), "Improved Testing and Specification of Smooth Transition Regression Models," in *Nonlinear Time Series Analysis of Economic and Financial Data*, ed. Rothman, P., Dordrecht: Kluwer Academic Publishers, pp. 289–320.

— (2001), "Testing Nonlinearity: Decision Rules for Selecting between Logistic and Exponential STAR Models," *Spanish Economic Review*, 3, 193–209.

Fama, E. F. (1965), "The Behavior of Stock Prices," *Journal of Business*, 38, 34–105.

Flachaire, E. (2005), "Bootstrapping Heteroskedastic Regression Models: Wild Bootstrap vs. Pairs Bootstrap," *Computational Statistics and Data Analysis*, 49, 361–376.

Frankel, J. (1986), "International Capital Mobility and Crowding-Out in the U.S. economy: Imperfect Integration of Financial Markets or of Goods Markets?" in *How Open Is the U.S. Economy?*, ed. Hafer, R. W., Massachusetts: Lexington, pp. 33–67.

Frankel, J. and Froot, K. (1990), "Chartists, Fundamentalists, and Trading in the Foreign Exchange Market," *American Economic Review*, 80, 181–85.

Frankel, J. and Rose, A. (1996), "A Panel Project on Purchasing Power Parity: Mean Reversion Within and between Countries," *Journal of International Economics*, 40, 209–224.

Frankel, J. A. (1990), "Zen and the Art of Modern Macroeconomics: A Commentary," in *Monetary Policy for a Volatile Global Economy*, eds. Haraf, W. S. and Willett, T. D., Washington, D.C.: AEI Press, pp. 117–23.

Froot, K. and Rogoff, K. S. (1995), "Perspectives on PPP and Long-Run Real Exchange Rates," in *The Handbook of International Economics*, eds. Grossman, G. and Rogoff, K. S., North Holland Amsterdam: Elsevier, pp. 1647–1688.

Gallagher, L. A. and Taylor, M. P. (2001), "Risky Arbitrage, Limits of Arbitrage, and Nonlinear Adjustment in the Dividend–Price Ratio," *Economic Inquiry*, 39, 524–36.

Gallant, A., Rossi, P., and Tauchen, G. (1993), "Nonlinear Dynamic Structures," *Econometrica*, 61, 871–908.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Gonçalves, S. and Kilian, L. (2004), "Bootstrapping Autoregressions With Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89–120.

— (2007), "Asymptotic and Bootstrap Inference for AR(∞) Processes With Conditional Heteroskedasticity." *Econometric Reviews*, 26, 609–641.

González, A. and Teräsvirta, T. (2006), "Simulation-based Finite Sample Linearity Test Against Smooth Transition Models," *Oxford Bulletin of Economics and Statistics*, 68, 797–812.

Granger, C. W. J. and Lee, T.-H. (1999), "The Effect of Aggregation on Nonlinearity," *Econometric Reviews*, 18, 259–269.

Granger, C. W. J. and Teräsvirta, T. (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press.

Haggan, V. and Ozaki, T. (1981), "Modeling Nonlinear Time Vibrations Using an Amplitude–Dependent Autoregressive Time Series Model," *Biometrika*, 68, 189–196.

Hamilton, J. (1994), *Time Series Analysis*, Princeton, NJ: Princeton University Press.

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

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997), "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 13, 281–291.

— (1998), "Tests for Forecast Encompassing," *Journal of Business & Economic Statistics*, 16, 254–59.

Head, A., Mattina, T., and Smith, G. (2004), "Real Exchange Rates, Preferences, and Incomplete Markets: Evidence, 1961-2001," *Canadian Journal of Economics*, 37, 782–801.

Hegwood, N. and Papell, D. (2002), "Purchasing Power Parity Under The Gold Standard," *Southern Economic Journal*, 69, 72–91.

Hinkley, D. V. (1977), "Jackknifing in Unbalanced Situations," *Technometrics*, 19, 285–292.

Horn, S. D., Horn, R. A., and Duncan, D. B. (1975), "Estimating Heteroskedastic Variances in Linear Models," *Journal of the American Statistical Association*, 70, 380–385.

Hsieh, D. A. (1992), "A Nonlinear Stochastic Rational Expectations Model of Exchange Rates," *Journal of International Money and Finance*, 235–250.

Hurn, S. and Becker, R. (2007), "Testing for Nonlinearity in Mean in the Presence of Heteroskedasticity," Working Paper Series 8, NCER.

Inoue, A. and Kilian, L. (2005), "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews*, 23, 371–402.

Jacks, D. S., Meissner, C. M., and Novy, D. (2008), "Trade Costs, 1870-2000," *American Economic Review*, 98, 529–34.

Jansen, E. S. and Teräsvirta, T. (1996), "Testing Parameter Constancy and Super Exogeneity in Econometric Equations," *Oxford Bulletin of Economics and Statistics*, 58, 735–63.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–80.

Johnston, L. D. and Williamson, S. H. (2008), "What Was the U.S. GDP Then?" MeasuringWorth.

Kapetanios, G. and Shin, Y. (2008), "GLS detrending-based unit root tests in nonlinear STAR and SETAR models," *Economics Letters*, 100, 377–380.

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

Kehoe, P. and Perri, F. (2002), "International Business Cycles With Endogenous Incomplete Markets," *Econometrica*, 70, 907–928.

Kilian, L. (1999), "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?" *Journal of Applied Econometrics*, 14, 491–510.

Kilian, L. and Taylor, M. P. (2001), "Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rates," Working Paper Series 088, European Central Bank.

— (2003), "Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, 60, 85–107.

Kiliç, R. (2004), "Linearity tests and stationarity," *Econometrics Journal*, 7, 55–62.

King, R., Plosser, C., and Rebelo, S. (1988), "Production, Growth and Business Cycles, I. The Basic Neoclassical Model," *Journal of Monetary Economcis*, 21, 195–232.

Koedijk, K., Nissen, F., and Schotman, P. (1996), "PPP and Real Expenditure: A Multicountry Panel Investigation," Limburg University, unpublished paper.

Kollmann, R. (1995), "Consumption, Real Exchange Rates and the Structure of International Asset markets," *Journal of International Money and Finance*, 14, 191–211.

Koop, G., Pesaran, M. H., and Potter, S. (1996), "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119–147.

Kreiss, J.-P. (1997), "Asymptotic Properties of Residual Bootstrap for Autoregressions," manuscript, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.

Krugman, P. R. (1991), "Target Zones and Exchange Rate Dynamics," *Quarterly Journal of Economics*, 106, 669–682.

Liu, R. (1988), "Bootstrap Procedure Under Some Non i.i.d. Models," *Annals of Statistics*, 16, 1696–1708.

Long, J. S. and Ervin, L. H. (2000), "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model," *The American Statistician*, 54, 217–224.

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

— (2008), "Real Exchange Rates Over the Past Two Centuries: How Important is the Harrod-Balassa-Samuelson Effect?" *Economic Journal*, 118, 1742–1763.

Lucas, R. (1982), "Interest Rates and Currency Price in a Two-Country World," *Journal of Monetary Economics*, 10, 335–359.

Lundbergh, S. and Teräsvirta, T. (1998), "Modelling Economic High Frequency Time Series With STAR–STGARCH Models," Stockholm School of Economics, Working Paper Series in Economics and Finance No. 291.

— (2002), "Forecasting with Smooth Transition Autoregressive Models," in *A companion to economic forecasting*, eds. Clements, M. P. and Hendry, D. F., Oxford: Blackwell, p. .

Lundbergh, S. and Teräsvirta, T. (2006), "A Time Series Model for an Exchange Rate in a Target Zone With Applications," *Journal of Econometrics*, 131, 579–609.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988), "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491499.

MacKinnon, J. G. and White, H. (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators With Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305–325.

Mammen, E. (1993), "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *Annals of Statistics*, 21, 255–285.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

Mark, N. C. (1990), "Real and Nominal Exchange Rates in the Long Run," *Journal of International Economics*, 28, 115–136.

McCracken, M. W. (2004), "Asymptotics for Out-of-Sample Tests of Causality," Manuscript, University of Missouri.

McMillan, D. G. and Speight, A. E. H. (2002), "Non-Linear Dynamics in High Frequency Intra-Day Financial Data: Evidence for the UK Long Gilt Futures Market," *Journal of Futures Markets*, 22, 1037–1057.

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

Mitchell, B. R. (2008a), *International Historical Statistics: Americas 1750-2005*, New York: Palgrave Macmillan.

— (2008b), *International Historical Statistics: Europe 1750-2005*, New York: Palgrave Macmillan.

Mohammed, S. I. S. and Williamson, J. G. (2004), "Freight rates and productivity gains in British tramp shipping 1869-1950," *Explorations in Economic History*, 41, 172–203.

Monoyios, M. and Sarno, L. (2002), "Mean Reversion in Stock Index Futures Markets: A nonlinear analysis," *Journal of Futures Markets*, 22, 285–314.

Mussa, M. (1986), "Nominal Exchange Rate Regimes and the Behaviour of Real Exchange Rates: Evidence and Implications," *Carnegie–Rochester Conference Series on Public Policy*, 25, 117–214.

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Novy, D. (2008), "Gravity Redux : Measuring International Trade Costs with Panel Data," The Warwick Economics Research Paper Series (TWERPS) 861, University of Warwick, Department of Economics.

Obstfeld, M. and Rogoff, K. S. (2000), "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" NBER Working Papers, No. 7777.

Obstfeld, M. and Taylor, A. M. (1997), "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited," *Journal of the Japanese and International Economies*, 11, 441–479.

O'Connell, P. G. J. and Wei, S. (2002), "The Bigger They Are, the Harder They Fall: Retail Price Differences Across U.S. Cities," *Journal of International Economics*, 56, 21–53.

Officer, L. H. (2008), "What Was the U.K. GDP Then?" MeasuringWorth.

Ozaki, T. (1978), "Non–Linear Models for Non-Linear Random Vibrations," Technical Report, Department of Mathematics, UMIST.

Park, J. (1992), "Canonical Cointegrating Regressions," *Econometrica*, 60, 119–143.

Pavlidis, E. G., Paya, I., and Peel, D. A. (2007), "Linearity Testing in the Presence of Heteroskedasticity," in *ESRC Seminar Series: Nonlinear Economics and Finance Research Community*, Brunel.

— (2009a), "The Econometrics of Exchange Rates," in *Palgrave Handbooks of Econometrics*, eds. Mills, T. and Patterson, K., London: Palgrave, vol. 2, p. ., part 5.4.

— (2009b), "Specifying Smooth Transition Regression Models in the Presence of Conditional Heteroskedasticity of Unknown Form," Working Papers 005913, Lancaster University Management School, Economics Department.

Paya, I. and Peel, D. A. (2005), "The Process Followed by PPP Data. On the Properties of Linearity Tests," *Applied Economics*, 37, 2515–2522.

— (2006a), "A New Analysis of the Determinants of the Real Dollar-Sterling Exchange Rate: 1871–1994," *Journal of Money, Credit and Banking*, 38, 1971–1990.

— (2006b), "Temporal Aggregation of an ESTAR Process: Some Implications for Purchasing Power Parity Adjustment," *Journal of Applied Econometrics*, 21, 655–668.

Paya, I., Venetis, I. A., and Peel, D. A. (2003), "Further Evidence on PPP Adjusment Speeds: The Case of Effective Real Exchange Rates and the EMS," *Oxford Bulletin of Economics and Statistics*, 65, 421–438.

Phillips, P. and Ouliaris, S. (1990), "Asymptotic Proporties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165–193.

Pippenger, M. and Goering, G. (1993), "A Note on the Empirical Power of Unit Root Tests Under Threshold Processes," *Oxford Bulletin of Economics and Statistics*, 55, 473–81.

Potter, S. (2000), "Nonlinear Impulse Response Functions," *Journal of Economic Dynamics and Control*, 24, 1425–1446.

Rajnish, Mehra, and Prescott, E. (1985), "The Equity Premium: A puzzle," *Journal of Monetary Economics*, 15, 145–161.

Rapach, D. E. and Wohar, M. E. (2006), "The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior," *International Journal of Forecasting*, 22, 341–361.

Rogoff, K. S. (1996), "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34, 647–668.

Rossi, B. (2005), "Testing Long-Horizon Predictive Ability With High Persistence, and the Meese-Rogoff Puzzle," *International Economic Review*, 46, 61–92.

Samuelson, P. (1964), "Theoretical Notes on Trade Problems," *Review of Economics and Statistics*, 46, 145–154.

Sarantis, N. (1999), "Modeling non-linearities in real effective exchange rates," *Journal of International Money and Finance*, 18, 27–45.

Sarno, L. and Taylor, M. P. (2002), *The Economics of Exchange Rates*, Cambridge England and New York: Cambridge University Press.

Selaive, J. and Tuesta, V. (2006), "The Consumption-Real Exchange Rate Anomaly: Non-Traded Goods, Incomplete Markets and Distribution Services," Working Papers Central Bank of Chile 359, Central Bank of Chile.

Sensier, M., Osborn, D. R., and Ocal, N. (2002), "Asymmetric Interest Rate Effects for the UK Real Economy," *Oxford Bulletin of Economics and Statistics*, 64, 315–39.

Sercu, P. and Uppal, R. (2000), *Exchange Rate Volatility, Trade, and Capital Flows Under Alternative Exchange Rate Regimes*, Cambridge, UK: Cambridge University Press.

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

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, eds. Cox, D. R., Hinkley, D. V., and Barndorff-Nielsen, O. E., London: Chapman & Hall, pp. 1–67.

Shleifer, A. and Vishny, R. (1997), "The Limits of Arbitrage," *Journal of Finance*, 52, 35–55.

Siddique, A. and Sweeney, R. J. (1998), "Forecasting real exchange rates1," *Journal of International Money and Finance*, 17, 63–70.

Skalin, J. and Teräsvirta, T. (1999), "Another Look at Swedish Business Cycles, 1861-1988," *Journal of Applied Econometrics*, 14, 359–78.

Stein, J., Allen, P., and Associates (1995), *Fundamental Determinants of Exchange Rates*, Oxford University Press.

Stock, J. H. and Watson, M. W. (1999), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger*, eds. Engle, R. F. and White, H., Oxford: Oxford University Press, p. .

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

Taylor, M. P. (1988), "An Epirical Examination of Long Run Purchasing Power Parity Using Cointegration Techniques," *Applied Economics*, 20, 1369–1381.

— (2006), “Real Exchange Rates and Purchasing Power Parity: Mean-reversion in Economic Thought,” *Applied Financial Economics*, 16, 1–17.

Taylor, M. P. and Peel, D. A. (2000), “Nonlinear adjustment, Long–Run Equilibrium and Exchange Rate Fundamentals,” *Journal of International Money and Finance*, 19, 33–53.

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

Taylor, N., van Dijk, D., Franses, P. H., and Lucas, A. (2000), “SETS, Arbitrage Activity, and Stock Price Dynamics,” *Journal of Banking and Finance*, 24, 1289–1306.

Taylor, S. (1986), *Modelling Financial Time Series*, .: John Wiley Chichester, Chichester.

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

— (2006), “Forecasting Economic Variables with Nonlinear Models,” in *Handbook of Economic Forecasting*, eds. Elliott, G., Granger, C. W. J., and Timmermann, A., London: North Holland, vol. 1, p. .

Teräsvirta, T., Lin, C.-F., and Granger, C. W. J. (1993), “Power of the Neural Network Linearity Test,” *Journal of Time Series Analysis*, 14, 209–220.

Tong, H. (1983), *Threshold Models in Non-Linear Time Series Analysis*, New York: Springer-Verlag.

Tsay, R. S. (1989), “Testing and Modeling Threshold Autoregressive Processes,” *Journal of the American Statistical Association*, 84, 231–240.

van Dijk, D. and Franses, P. H. (2003), "Selecting a Nonlinear Time Series Model using Weighted Tests of Equal Forecast Accuracy," *Oxford Bulletin of Economics and Statistics*, 65, 727–743.

van Dijk, D., Franses, P. H., and Boswijk, P. (2007), "Absorption of Shocks in Nonlinear Autoregressive Models," *Computational Statistics and Data Analysis*, 51, 4206–4226.

van Dijk, D., Franses, P. H., Clements, M. P., and Smith, J. (2003), "On SETAR non-linearity and forecasting," *Journal of Forecasting*, 22, 359–375.

van Dijk, D., Teräsvirta, T., and Frances, P. (2002), "Smooth Transition Autoregressive Models – A Survey of Recent Developments," *Econometrics Reviews*, 21, 1–47.

Venetis, I. A., Paya, I., and Peel, D. A. (2007), "Deterministic Impulse Response in a Nonlinear Model. An Analytic Expression," *Economics Letters*, 95, 315–319.

Vogelsang, T. J. (1998), "Trend function hypothesis testing in the presence of serial correlation," *Econometrica*, 66, 123–48.

West, K. D. (2006), "Forecast Evaluation," in *Handbook of Economic Forecasting*, eds. Elliott, G., Granger, C. W. J., and Timmermann, A., London: North Holland, vol. 1, p. .

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–38.

Wong, C.-S. and Li, W.-K. (1997), "Testing for Threshold Autoregression With Conditional Heteroscedasticity," *Biometrika*, 84, 407–418.

Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis (with discussion)," *Annals of Statististics*, 14, 1261–1350.

Yadav, P. K., Pope, P., and Paudyal, K. (1994), "Threshold Autoregressive Modeling in Finance: The Price Differences of Equivalent Assets," *Mathematical Finance*, 4, 205–221.