Lancaster University Management School
Working Paper
2009/009

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

E Pavlidis, Ivan Paya and David Peel

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

The LUMS Working Papers series can be accessed at http://www.lums.lancs.ac.uk/publications/
LUMS home page: http://www.lums.lancs.ac.uk/

Specifying Smooth Transition Regression Models in the Presence of Conditional Heteroskedasticity of Unknown Form*

Efthymios G. Pavlidis Ivan Paya† David A. Peel
Lancaster University Management School, Lancaster, LA1 4YX, UK

ABSTRACT

The specification of Smooth Transition Regression models consists of a sequence of tests, which are typically based on the assumption of *i.i.d.* errors. In this paper we examine the impact of conditional heteroskedasticity and investigate the performance of several heteroskedasticity robust versions. Simulation evidence indicates that conventional tests can frequently result in finding spurious nonlinearity. Conversely, when the true process is nonlinear in mean the tests appear to have low size adjusted power and can lead to the selection of misspecified models. The above deficiencies also hold for tests based on Heteroskedasticity Consistent Covariance Matrix Estimators but not for the Fixed Design Wild Bootstrap. We highlight the importance of robust inference through empirical applications.

KEY WORDS: Time Series, Robust Linearity Test, Heteroskedasticity Consistent Covariance Matrix Estimator, Wild Bootstrap, Monte Carlo Simulation
JEL Classification: C15, C52

*The authors are grateful to participants of the second meeting of the ESRC Research Seminar series on "Nonlinear Economics and Finance Research Community" held in Brunel and participants of the 16^{th} Annual Symposium of the Society for Nonlinear Dynamics and Econometrics. Ivan Paya acknowledges financial support from the Spanish Ministerio de Educacion y Ciencia Research Project SEJ2005- 02829/ECON.

†Correspondence to: Ivan Paya, Lancaster University Management School, Lancaster, LA1 4YX, UK. E-mail: i.paya@lancaster.ac.uk

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 Granger and Teräsvirta (1993) and Teräsvirta (1994). By allowing regime dependent behavior 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 modeling 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 LM 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 (1993, 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 modeling procedure. However, such a modeling cycle may often lead to the misspecification of the conditional mean.

In this article, 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 utilize 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, 2007). HCCMEs are typically employed by researchers due to their simple implementation and their little computational cost (Long and Ervin 2000) compared to bootstrap methods. Although we focus on the Generalized 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 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 paper is organized as follows. Section 2 outlines the basic STAR representation, which facilitates the analysis of testing linearity against STAR nonlinearity in Section 3. Dealing with conditional heteroskedasticity of unknown form using HCCMEs and the Fixed Design Wild Bootstrap is discussed in Section 3.1. The next section investigates the finite sample performance of the tests through Monte Carlo simulations. Section 5 presents an empirical application on empirical data. Finally, the last section concludes.

2 THE SMOOTH TRANSITION REGRESSION MODEL

The basic STAR model representation for a univariate time series y_t is given by

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

or equivalently

$$y_t = \boldsymbol{\pi}_1' \boldsymbol{x}_t + \boldsymbol{\pi}_2' \boldsymbol{x}_t F(s_t; \gamma, c) + \epsilon_t, \qquad t = 1, \ldots, T, \quad (2)$$

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) with exogenous regressors. Hence, our analysis can be generalized 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

$$F(s_t; \gamma, c) = 1 - \exp\left(-\gamma \left(s_t - c\right)^2\right), \qquad \gamma > 0, \quad (3)$$

while for the Logistic STAR (LSTAR),

$$F(s_t;\gamma,c) = [1+\exp(-\gamma(s_t-c))]^{-1}, \qquad \gamma > 0, \tag{4}$$

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 $(y_{t-d}-c)$ and admits the limits

$$F(\cdot) \;\rightarrow\; 1 \qquad \text{as} \qquad |s_{t-d}-c| \rightarrow +\infty, \tag{5}$$

$$F(\cdot) \;\rightarrow\; 0 \qquad \text{as} \qquad |s_{t-d}-c| \rightarrow 0. \tag{6}$$

While the Logistic transition function is asymmetric around $(y_{t-d}-c)$ and admits the limits

$$F(\cdot) \;\rightarrow\; 1 \qquad \text{as} \qquad (s_{t-d}-c) \rightarrow +\infty, \tag{7}$$

$$F(\cdot) \;\rightarrow\; 0 \qquad \text{as} \qquad (s_{t-d}-c) \rightarrow -\infty. \tag{8}$$

Figure 1 illustrates the two transition functions. 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. 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.

FIGURE 1

The properties of STR and STAR models are very appealing in modeling nonlinear economic and financial time series. For example, the fact that macroeconomic time series as well as their relationships may be characterized by asymmetries associated with the stages of the business cycle (see, e.g., Skalin and Teräsvirta 1999; Sensier et al. 2002; Deschamps 2008) makes LSTR 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) motivating the use of ESTR models (e.g., Michael et al. 1997; Gallagher and Taylor 2001; McMillan and Speight 2002).

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 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) gives rise to an nuisance parameter problem (Davies 1977, 1987). The null hypothesis of linearity corresponds to both H_0: $\boldsymbol{\pi}_2' = 0$ 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 $\boldsymbol{\pi}_2'$ 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 $\gamma = 0$.[1] This re-parameterization resolves the identification problem since it does not involve nuisance parameters. The auxiliary regression is given by

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \boldsymbol{x}_t s_t + \boldsymbol{\delta}_2' \boldsymbol{x}_t s_t^2 + u_t, \tag{9}$$

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

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \tilde{\boldsymbol{x}}_t s_t + \boldsymbol{\delta}_2' \tilde{\boldsymbol{x}}_t s_t^2 + u_t, \tag{10}$$

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

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

$$y_t = \boldsymbol{\delta}_0'\boldsymbol{x}_t + \boldsymbol{\delta}_1'\boldsymbol{x}_t s_t + \boldsymbol{\delta}_2'\boldsymbol{x}_t s_t^2 + \boldsymbol{\delta}_3'\boldsymbol{x}_t s_t^3 + u_t. \tag{11}$$

Teräsvirta (1994) proposes a modeling procedure based on Equation (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, 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.

While Teräsvirta (1994) uses a third order Taylor expansion of the logistic function and a first order 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

$$y_t = \boldsymbol{\delta}_0' \boldsymbol{x}_t + \boldsymbol{\delta}_1' \boldsymbol{x}_t s_t + \boldsymbol{\delta}_2' \boldsymbol{x}_t s_t^2 + \boldsymbol{\delta}_3' \boldsymbol{x}_t s_t^3 + \boldsymbol{\delta}_4' \boldsymbol{x}_t s_t^4 + u_t. \tag{12}$$

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 of LSTAR nonlinearity, $H_0^L : \boldsymbol{\delta}_2' = \boldsymbol{\delta}_4' = 0$, with an F test, (F_L).
2. Test the null of ESTAR nonlinearity, $H_0^E : \boldsymbol{\delta}_1' = \boldsymbol{\delta}_3' = 0$, 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) is independent, identically and normally distributed. However, the assumption of constant conditional variance may be too strict when it comes to empirical applications.

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 2006; 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), Taylor and Kilian (2003) and Paya et al. (2003) show that ESTAR models can capture the behavior 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 modeling 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 non parametric approach to deal with conditional heteroskedasticity of unknown form in Equation (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).

3.2 Hypothesis Testing

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

$$y_t = \boldsymbol{\delta}'\boldsymbol{z}_t + u_t, \tag{13}$$

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

$$W = \left(\boldsymbol{R}\widehat{\boldsymbol{\delta}}\right)' \left(\boldsymbol{R}\widehat{\boldsymbol{\Psi}}\boldsymbol{R}'\right)^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\delta}}\right), \tag{14}$$

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 $\widehat{\Psi}$ is required when drawing inferences. Assuming that the residuals, u_t, are independent, identically and normally distributed with variance σ_u^2 yields

$$\text{LS} : \widehat{\boldsymbol{\Omega}} = \widehat{\sigma}_u^2 \boldsymbol{I}, \tag{15}$$

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). In this case, HCCMEs are usually employed by researchers. Eicker (1963) and White (1980) propose the following heteroskedasticity consistent estimator

$$\text{HC0} : \ \widehat{\boldsymbol{\Omega}} = \text{diag}(\widehat{u}_t^2), \tag{16}$$

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

$$\text{HC2}: \widehat{\boldsymbol{\Omega}} = \operatorname{diag}\left(\frac{\widehat{u}_t^2}{1-h_{tt}}\right), \tag{17}$$

$$\text{HC3}: \widehat{\boldsymbol{\Omega}} = \operatorname{diag}\left(\frac{\widehat{u}_t^2}{(1-h_{tt})^2}\right), \tag{18}$$

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. Consequently, tests based on asymptotic theory may lead to incorrect inferences in finite samples.

Rather than relying on hypothesis tests based on asymptotic theory, bootstrap methods can be employed for conducting statistical inference. The rationale in bootstrap methods is to approximate the finite sample distribution of the test statistic under the null by simulation. In small samples, bootstrap tests may lead to a significant improvement in term of the error in rejection probabilities (see, e.g, MacKinnon 2002, 2006).

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, 2007) 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 (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.

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 ESTAR(1) conditional mean equation examined by Escribano and Jordá (2001)

$$y_t = \pi_{1,1} y_{t-1} + \pi_{2,1} y_{t-1}[1 - \exp(-\gamma y_{t-d}^2)] + \epsilon_t, \qquad t = 1, \ldots, T, \tag{19}$$

where $\pi_{1,1} = 0.3$ and $\pi_{2,1} = -0.9$. 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,

$$\epsilon_t = e_t h_t^{1/2}, \quad h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}, \quad e_t \sim \mathcal{N}(0,1) \tag{20}$$

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:

$$\begin{aligned} \epsilon_t &= e_t h_t^{1/2}, \quad \ln(h_t) = -0.23 + 0.9\ln(h_{t-1}) + 0.25\left(e_{t-1}^2 - 0.3e_{t-1}\right), \\ e_t &\sim \mathcal{N}(0,1). \end{aligned} \tag{21}$$

AGARCH:

$$\begin{aligned} \epsilon_t &= e_t h_t^{1/2}, \quad h_t = 0.0216 + 0.6896 h_{t-1} + 0.3174\left(\epsilon_{t-1} - 0.1108\right)^2, \\ e_t &\sim \mathcal{N}(0,1). \end{aligned} \tag{22}$$

GJR GARCH:

$$\begin{aligned} \epsilon_t &= e_t h_t^{1/2}, \quad h_t = 0.005 + 0.7 h_{t-1} + 0.28 \left(\epsilon_{t-1}^2 - 0.23 \epsilon_{t-1} \right), \\ e_t &\sim \mathcal{N}(0, 1). \end{aligned} \tag{23}$$

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" characterizing 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 Deutsche mark against the pound sterling.

$$\begin{aligned} \epsilon_t &= e_t \exp(h_t), \quad h_t = 0.951 h_{t-1} + 0.5 e_t, \\ (\epsilon_t, e_t) &\sim \mathcal{N}(0, \mathrm{diag}(0.18, 1)). \end{aligned} \tag{24}$$

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.[2] The first 100 observations are discarded to avoid initialization effects.

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 (19) becomes an AR(1) model.

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

Tables 1 and 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 labeled LS, HC0, HC2 and HC3, and WB, respectively. In addition, Figure 2 provides a visual view of the error of rejection probability (the difference between the empirical size and the nominal level of a test) for stationary GARCH processes.

FIGURE 2

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 Terasvirta (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). 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 favor either alternative (ESTAR and LSTAR), which is also true for the remaining inference techniques.

TABLE 1

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 HCCME, 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

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 $\gamma = 1$ and the transition variable equal to y_{t-1}. The details for the simulation procedure are the same as before. Tables 3 and 4 report the results.

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.

TABLE 3

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 of Shephard (1996) 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.

TABLE 4

4.3 Nonlinear Model Specification

So far we 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 minimizes 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 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 5

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 behavior 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 modeling 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.

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 (1963, p. 418) 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...*". A similar phenomenon is observed for other asset returns, such as exchange rates (Baillie and Bollerslev 1991, 2002).

However, time varying volatility is not constraint to high frequency data. The findings of several empirical studies suggest that the volatity 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 2006; Paya and Peel 2006a). To this end, we employ the Lothian and Taylor (1996) two century data 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 Taylor and Kilian (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 behavior 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 1$^{\text{st}}$, 1991 to the December 31$^{\text{st}}$, 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 1$^{\text{st}}$, 1988 to December 31$^{\text{st}}$, 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

$$b_{t,k} = \ln\left(\frac{F_{t,k}}{P_t}\right), \tag{25}$$

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.

FIGURE 3

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

$$\widehat{\epsilon}_t^2 = \mu + \sum_{i=1}^{q} a_i \widehat{\epsilon}_{t-i}^2 + v_t, \tag{26}$$

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 and the sterling dollar real exchange rate. For the latter two series we follow Monoyios and Sarno (2002) and Lothian and Taylor (2006) and set the lag length to five and two, respectively. 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 6

Not surprisingly, Table 6 shows that the null hypothesis can be rejected at all conventional levels of significance for all high frequency series. Note, however, that like nonlinear in mean tests have power in detecting ARCH effects, ARCH tests have power in detecting nonlinearity in mean (Lee et al. 1993; Blake and Kapetanios 2007). Therefore, rejection of the null hypothesis may be due to the presence of STAR type nonlinearity.

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 7 reports the P values for the null of linearity corresponding to each transition variable and the selected model.

TABLE 7

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 characterized 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.

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 paper 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. Our findings 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.

REFERENCES

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.

Bera, A. K. and Higgins, M. L. (1993), "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys*, 7, 305–366.

— (1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence," *Journal of Business and Economic Statistics*, 15, 43–50.

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

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.

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.

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.

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.

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.

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

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

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.

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 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.

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.

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.

Hansen, B. E. (1996), "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–30.

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.

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.

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.

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

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.

— (2006), "Real Exchange Rates Over the Past Two Centuries: How Important Is the Harrod–Balassa–Samuelson Effect?" The Warwick Economics Research Paper Series (TWERPS) 768, University of Warwick, Department of Economics.

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.

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 Model," *Biometrika*, 75, 491–499.

MacKinnon, J. G. (2002), "Bootstrap Inference in Econometrics," *Canadian Journal of Economics*, 35, 615–645.

— (2006), "Bootstrap Methods in Econometrics," *The Economic Record*, 82, S2–S18.

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.

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

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.

Paya, I. and Peel, D. A. (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.

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.

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.

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

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

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.

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.

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

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

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.

7 TABLES & FIGURES

Table 1: Empirical Size of Wald F Tests

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

			H_0: Linearity					ESTAR selection				
T	α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	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
250	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
500	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
	0.05	0.94	0.11	0.33	0.19	0.10	0.04	0.05	0.20	0.11	0.05	0.02
1000	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 (H_0: Linearity) 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: 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
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:	$y_t = 0.3y_{t-1} + \epsilon_t,\ \epsilon_t = e_t \exp(h_t)$, $h_t = 0.951h_{t-1} + 0.5e_t,\ (\epsilon_t, e_t) \sim \mathcal{N}(0, \mathrm{diag}(0.18, 1))$.									
	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB

Continued on Next Page…

Table 2: Empirical Size of Wald F Tests(Cont'd.)

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: The table reports the empirical size of the LS, HC0, HC2, HC3 and the WB linearity tests (H_0: Linearity) 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 3: Empirical Size Adjusted Power of Wald F Tests

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

			H_0: Linearity					ESTAR selection				
T	α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	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
250	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
500	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
	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
1000	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

Continued on Next Page...

Table 3: Empirical Size Adjusted Power of Wald F Tests(Cont'd.)

			H_0: Linearity					ESTAR selection				
T	α	β	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
	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, and HC3 linearity tests (H_0: Linearity) and the power of the WB tests. It also shows the percentage of times an ESTAR model is selected rather than an LSTAR (ESTAR selection). The nominal significance level is 5%.

Table 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: $y_t = 0.3y_{t-1} - 0.9y_{t-1}[1 - \exp(-y_{t-1}^2)] + \epsilon_t$, $\epsilon_t = e_t h_t^{1/2}$
$h_t = 0.0216 + 0.6896h_{t-1} + 0.3174\,(\epsilon_{t-1} - 0.1108)^2$, $e_t \sim \mathcal{N}(0,1)$.

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

Continued on Next Page...

Table 4: Empirical Size Adjusted Power of Wald F Tests(Cont'd.)

	$h_t = 0.005 + 0.7h_{t-1} + 0.28\left(\epsilon_{t-1}^2 - 0.23\epsilon_{t-1}\right)$, $e_t \sim \mathcal{N}(0,1)$.									
	H_0: Linearity					ESTAR selection				
T	LS	HC0	HC2	HC3	WB	LS	HC0	HC2	HC3	WB
100	0.12	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: $y_t = 0.3y_{t-1} - 0.9y_{t-1}[1 - \exp(-y_{t-1}^2)] + \epsilon_t$, $\epsilon_t = e_t \exp(h_t)$,
$h_t = 0.951h_{t-1} + 0.5e_t$, $(\epsilon_t, e_t) \sim \mathcal{N}(0, \mathrm{diag}(0.18, 1))$.

	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: The table reports the empirical size adjusted power of the LS, HC0, HC2, and HC3 linearity tests (H_0: Linearity) and the power of the WB tests. It also shows the percentage of times an ESTAR model is selected rather than an LSTAR (ESTAR selection). The nominal significance level is 5%.

Table 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

Continued on Next Page…

Table 5: Selection Frequencies of the Delay Parameter, d(Cont'd.)

T	delay	LS	HC0	HC2	HC3	WB
	$d=3$	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
	$d=3$	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

Continued on Next Page...

Table 5: Selection Frequencies of the Delay Parameter, d(Cont'd.)

T	delay	LS	HC0	HC2	HC3	WB
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} when the error term exhibits conditional heteroskedasticity. True delay parameters are in bold.

Table 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.

Table 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 STERLING	LS	**0.000**	0.000	0.000	0.004	LSTAR
	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 BASIS	LS	**0.000**	0.010	0.071	0.524	LSTAR
	HC0	**0.000**	0.003	0.000	0.156	LSTAR
	HC2	**0.000**	0.022	0.030	0.314	LSTAR
	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

Continued on Next Page…

Table 7: Application of Linearity Tests on Empirical Data(Cont'd.)

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%.

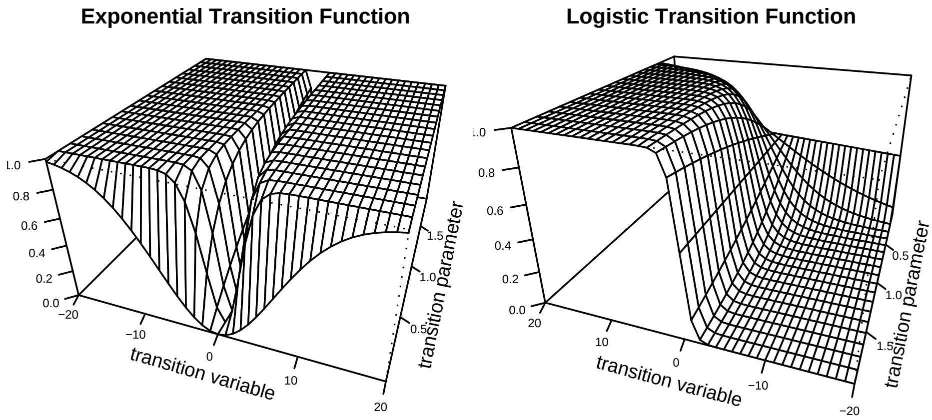

Figure 1: The Logistic and Exponential Transition Functions for $\gamma \in \{0.01, \ldots, 2\}$, $s_t \in \{-20, \ldots, 20\}$ and $c = 0$.

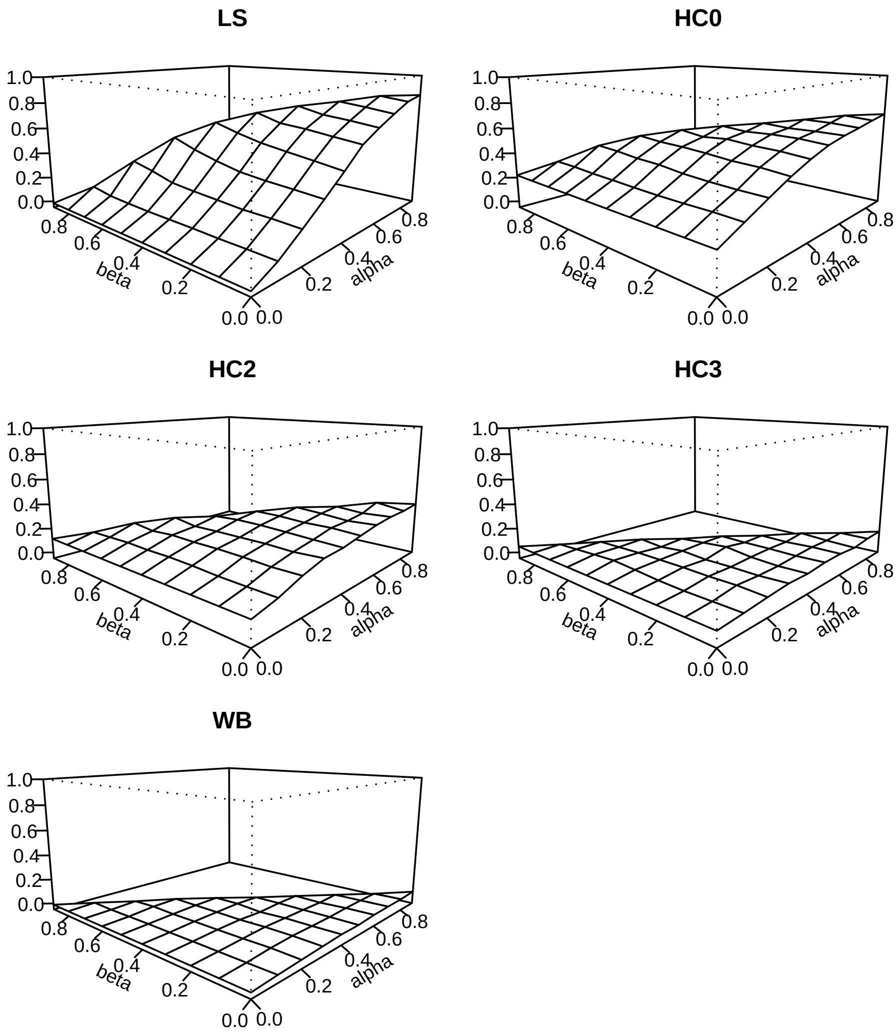

Figure 2: Error of rejection probability in LS, HC0, HC2, HC3 and WB linearity tests in the presence of conditional heteroskedasticity. The DGP is an AR(1)-GARCH(1,1) model. The AR coefficient $\phi = 0.3$, and the GARCH parameters $\alpha \in \{0, 0.1, \ldots, 0.8, 0.9\}$ and $\beta \in \{0, 0.1, \ldots, 0.8, 0.9\}$ satisfy $\alpha + \beta < 1$. The unconditional variance of the error process is set to unity ($\omega = 1 - \alpha - \beta$).

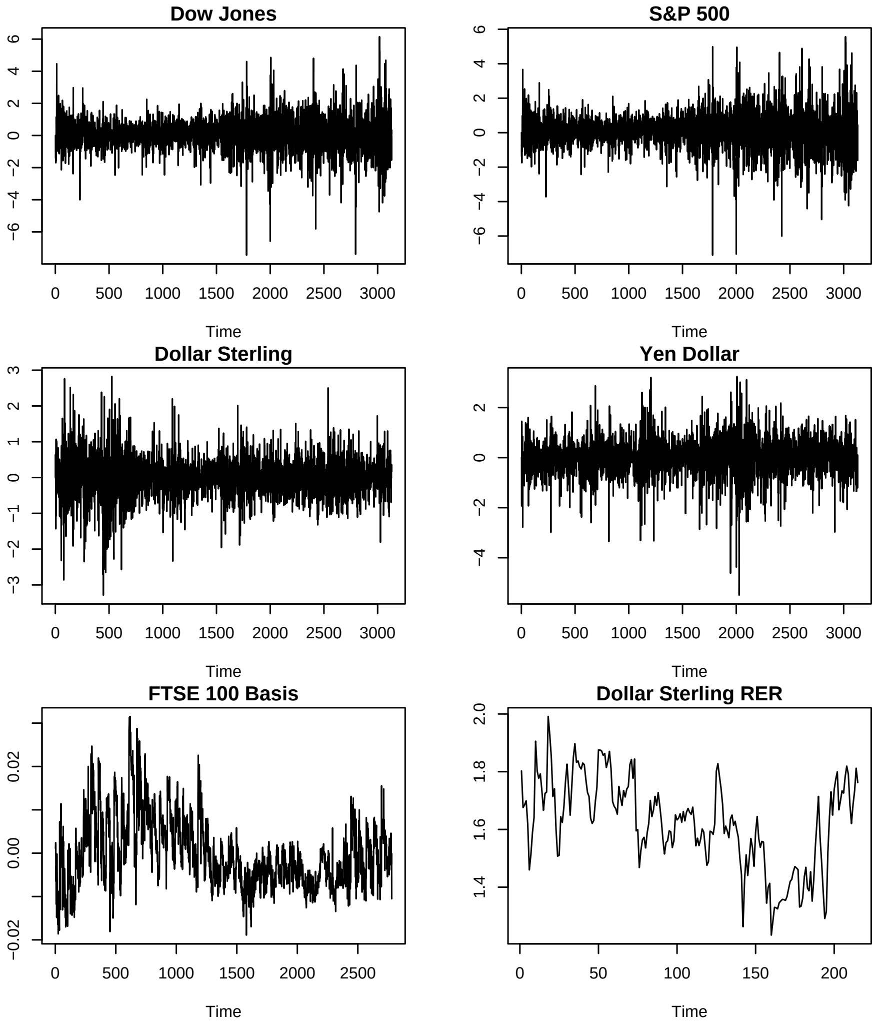

Figure 3: Time series plots of empirical data. Daily returns on the Dow Jones and the S&P 500 indices, and the yen-dollar and dollar-sterling nominal exchange rates cover the period January 2nd, 1991 to the December 31st, 2002. The basis of the FTSE 100 spans the period January 2nd, 1988 to December 31st, 1998, and the dollar-sterling real exchange rate (RER) the period 1791 to 2005.