Econometric Institute Research Report 9832/A

Nonlinearities and Outliers: Robust Specification of STAR Models

Álvaro Escribano
Departamento de Estadística y Econometría, Universidad Carlos III de Madrid

Philip Hans Franses
Econometric Institute, Erasmus University Rotterdam

Dick van Dijk*
Tinbergen Institute, Erasmus University Rotterdam

August 10, 1998

Abstract

Outliers and nonlinearity may easily be mistaken. This paper uses Monte Carlo methods to examine and compare the behavior of two competing specification procedures for Smooth Transition AutoRegressive [STAR] models under various different circumstances (linear and nonlinear data generating processes, with and without outlier contamination). The extensive simulation evidence demonstrates that the use of outlier-robust variants of the linearity tests which are involved leads to procedures with more desirable properties. An application to several real exchange rate series illustrates the potential usefulness of the robust specification procedures, especially in case one is not certain whether or not aberrant observations are present.

Keywords: Smooth transition autoregression, Lagrange Multiplier test, Specification procedure, Outliers.

JEL classification: C12, C22

*Correspondence to Dick van Dijk, Tinbergen Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `email: djvandijk@few.eur.nl`. Part of this research was done while the third author was visiting the Departamento de Estadística y Econometría at the Universidad Carlos III de Madrid (September 1997). Their hospitality and the stimulating environment is highly appreciated.

1 Introduction

Among the myriad of nonlinear time series models which have been proposed over the years, Smooth Transition AutoRegressive [STAR] models are enjoying a fair amount of popularity. These models have recently been applied to describe nonlinearities in the business cycle (Teräsvirta and Anderson 1992, Skalin and Teräsvirta 1996, Skalin and Teräsvirta 1998, Van Dijk and Franses 1998a), the term structure of interest rates (Anderson 1997, Van Dijk and Franses 1998b), money demand (Wolters, Teräsvirta and Lütkepohl 1996), and real exchange rates (Michael, Nobay and Peel 1997, Baum, Caglayan and Barkoulas 1998), among others.

It has become standard practice to use the procedure outlined by Teräsvirta (1994) to specify empirical STAR models (for example, the procedure is applied in all of the papers mentioned above). Recently, the properties of this procedure have been investigated in more detail, and some potential difficulties have been pointed out. First, it has been criticized by Escribano and Jordá (1997) for not being able to discriminate between various different forms of the STAR model which are used in practice. Escribano and Jordá (1997,1998) therefore suggest an alternative specification procedure with more desirable properties. Second, STAR models can be parameterized in such a way that they generate very asymmetric realizations, that is, the resultant data resemble time series with a few outliers. Van Dijk, Franses and Lucas (1998) examine the behavior of Lagrange Multiplier [LM] tests for STAR type nonlinearity developed by Luukkonen, Saikkonen and Teräsvirta (1988), which form an essential ingredient of the specification procedure of Teräsvirta (1994), in the presence of outliers. It is found that such aberrant observations can substantially distort the distributional properties of the test statistics. In particular, the tests become biased towards rejecting the correct null hypothesis of linearity. To overcome this, Van Dijk *et al.* (1998) develop robust variants of the linearity tests which are (more) resistant to the presence of outliers. The main advantage of this robust procedure is that it automatically guards the tests against outliers, and does not require a priori knowledge concerning their presence and timing.

In the present paper, we attempt to provide further insight into specifying STAR models by combining the results in Escribano and Jordá (1997,1998) with those in Van Dijk *et al.* (1998). In particular, the following three questions are addressed: 1) What happens to regular specification procedures for STAR models in case only outliers are present?, 2) What happens to outlier-robust specification procedures in case only nonlinearity is present?, and 3) What happens to regular and robust procedures if both nonlinearity and outliers occur? These questions are addressed in Sections 3 and 4. The analysis is preceded by a brief discussion of STAR models and the available specification procedures in Section 2. Section 5 illustrates the available and newly developed specification procedures by applying them to gold and silver prices and to several real exchange rate series. Finally, Section 6 summarizes our findings and discusses some practical guidelines.

2 The STAR model: tests and specification procedures

Consider the general STAR(p) model

$$y_t = \phi' y_t^{(p)} + F(y_{t-d}; \gamma, c)\theta' y_t^{(p)} + \varepsilon_t, \tag{1}$$

where $y_t^{(p)} = (1, \tilde{y}_t^{(p)})', \tilde{y}_t^{(p)} = (y_{t-1}, \ldots, y_{t-p})'$, $\phi = (\phi_0, \phi_1, \ldots, \phi_p)'$, $\theta = (\theta_0, \theta_1, \ldots, \theta_p)'$, and $\varepsilon_t \sim$ i.i.d.$N(0, \sigma^2)$. The transition function $F(y_{t-d}; \gamma, c)$ is usually taken to be either the logistic function[1]

$$F(y_{t-d}; \gamma, c) = (1 + \exp\{-\gamma(y_{t-d} - c)\})^{-1} - \frac{1}{2}, \qquad \gamma > 0, \tag{2}$$

or the exponential function[2]

$$F(y_{t-d}; \gamma, c) = 1 - \exp\{-\gamma(y_{t-d} - c)^2\}, \qquad \gamma > 0, \tag{3}$$

where y_{t-d} is called the transition variable and d the delay parameter. Model (1) with transition function (2) is called the Logistic STAR [LSTAR] model, while model (1) with transition function (3) is called the Exponential STAR [ESTAR] model.

It is easily understood that the LSTAR and ESTAR models imply quite different behavior for the series y_t. In the LSTAR model, the dynamics are different for small and large values of y_{t-d} (relative to the threshold c). The ESTAR model on the other hand

[1] The reason for subtracting $\frac{1}{2}$ in (2) is that it facilitates the derivation of the test statistics. It should be noted that in the Monte Carlo simulations in Section 4, this constant is not subtracted.

[2] Jansen and Teräsvirta (1996) argue that the exponential function suffers from the drawback that it does not nest a Threshold AutoRegressive [TAR] model as a limiting case (because when either $\gamma \to 0$ or $\gamma \to \infty$, the model collapses to a linear model). These authors propose an alternative function which does nest a three-regime TAR model as a limiting case. Because this does not affect the tests for linearity and the specification procedures for STAR models which are the main subject of this paper, we do not discuss this point any further.

implies similar dynamics for small and large values of y_{t-d}, while the dynamics are different for values of y_{t-d} close to and far from c.

The STAR model as given above can easily be extended to include exogenous variables, either as regressors or as transition variables or both, see Granger and Teräsvirta (1993), Teräsvirta (1998), and Escribano and Jordá (1998) for extensive discussions of the resulting class of Smooth Transition Regression [STR] models.

An obvious specification procedure for STAR models is to test for the presence of STAR-type nonlinearity first and next to decide between the logistic and exponential STAR models. These two elements are discussed in the following subsections.

2.1 Testing linearity

Luukkonen, Saikkonen and Teräsvirta (1988) [LST hereafter] consider testing the null hypothesis of linearity against the alternative of LSTAR nonlinearity. The null hypothesis, which might be expressed as $H_0 : \gamma = 0$ in (2), cannot be tested using standard techniques, because under the null, the parameters θ and c in (1) are not identified. LST suggest to circumvent this problem by replacing the transition function (2) by a third-order Taylor approximation around the null hypothesis[3], which after rearranging terms yields the auxiliary regression model[4]

$$y_t = \phi' y_t^{(p)} + \beta_1' \tilde{y}_t^{(p)} y_{t-d} + \beta_2' \tilde{y}_t^{(p)} y_{t-d}^2 + \beta_3' \tilde{y}_t^{(p)} y_{t-d}^3 + \eta_t, \tag{4}$$

where the $\beta_i = (\beta_{i1}, \ldots, \beta_{ip})'$, $i = 1, 2, 3$, are functions of the parameters θ, γ and c. The null hypothesis now becomes $H_0' : \beta_1 = \beta_2 = \beta_3 = 0$, which can be tested by a standard Lagrange Multiplier [LM] test in a straightforward manner. Under the null hypothesis of linearity, the test statistic has a χ^2 distribution with $3p$ degrees of freedom asymptotically. Following Escribano and Jordá (1997), we will denote this test statistic as NL3.

Granger and Teräsvirta (1993) suggest that linearity might be tested against an ESTAR alternative by replacing the function (2) with a first-order Taylor approximation

[3]That is, $F(y_{t-d}; \gamma, c) \approx \gamma \left. \frac{\partial F(y_{t-d};\gamma,c)}{\partial \gamma} \right|_{\gamma=0} + \frac{1}{6}\gamma^3 \left. \frac{\partial^3 F(y_{t-d};\gamma,c)}{\partial \gamma^3} \right|_{\gamma=0} = \frac{1}{4}\gamma(y_{t-d} - c) + \frac{1}{48}\gamma^3(y_{t-d} - c)^3$, where we have used the facts that $F(y_{t-d}; 0, c) = 0$ and that the second derivative of F with respect to γ evaluated at $\gamma = 0$ equals zero as well. Note that the same auxiliary regression is obtained for all transition functions F which share the property of the logistic function (2) that all even-ordered derivatives are equal to zero at $\gamma = 0$.

[4]It should be noted that if $d > p$ (and hence y_{t-d} is not included in $\tilde{y}_t^{(p)}$), additional terms $\beta_{10} y_{t-d}$, $\beta_{20} y_{t-d}^2$, and $\beta_{30} y_{t-d}^3$ should be included in (4). The test statistic should be adjusted to accommodate the fact that the parameters $\beta_{i0}, i = 1, 2, 3$ should also be equal to zero under the null hypothesis.

around $\gamma = 0$, which gives rise to an auxiliary model similar to (4), only without the term $\beta_3 \tilde{y}_t^{(p)} y_{t-d}^3$. Hence, Teräsvirta (1994) suggests that the LM-type test based upon (4) should have power against both LSTAR and ESTAR alternatives and might be used as a test against general STAR-type nonlinearity.

Escribano and Jordá (1997) argue that a first-order approximation for the exponential function is not sufficient to capture its distinguishing characteristics, in particular the two inflexion points of this function. Hence they conclude that a *second*-order Taylor approximation is necessary[5], yielding the auxiliary regression,

$$y_t = \phi' y_t^{(p)} + \beta_1' \tilde{y}_t^{(p)} y_{t-d} + \beta_2' \tilde{y}_t^{(p)} y_{t-d}^2 + \beta_3' \tilde{y}_t^{(p)} y_{t-d}^3 + \beta_4' \tilde{y}_t^{(p)} y_{t-d}^4 + \eta_t, \tag{5}$$

The null hypothesis to be tested now is $H_0' : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. The resulting LM-type test statistic, denoted NL4, has an asymptotic χ^2 distribution with $4p$ degrees of freedom under the null hypothesis.

2.2 Specification procedures

Note that in the derivation of the LM-type test statistics, the value of d which determines the transition variable has been assumed known[6]. In practice, the tests are calculated for different values of d (e.g., $d = 1, \ldots, p$), and the value of d for which the null hypothesis is rejected most convincingly is selected as the delay parameter.

Once linearity is rejected in favor of STAR-type nonlinearity by either NL3 or NL4, one has to decide between using (2) or (3) in the STAR model (1) (or similar functions which have the same properties). Teräsvirta (1994) suggests to use a decision rule based upon a sequence of tests nested within the null hypothesis corresponding to (4). In particular, he proposes to test the hypotheses

$$\begin{aligned} H_{03}: &\quad \beta_3 = 0, \\ H_{02}: &\quad \beta_2 = 0 |\ \beta_3 = 0, \\ H_{01}: &\quad \beta_1 = 0 |\ \beta_3 = \beta_2 = 0, \end{aligned}$$

by means of LM-type tests. Closer inspection of the expressions of the auxiliary parameters β_1, β_2 and β_3 in (4) in terms of parameters of the original STAR model reveals that (i) β_3

[5] That is, $F(y_{t-d}; \gamma, c) \approx \gamma \left. \frac{\partial F(y_{t-d};\gamma,c)}{\partial \gamma} \right|_{\gamma=0} + \frac{1}{2}\gamma^2 \left. \frac{\partial^2 F(y_{t-d};\gamma,c)}{\partial \gamma^2} \right|_{\gamma=0} = \gamma(y_{t-d} - c)^2 - \frac{1}{2}\gamma^2 (y_{t-d} - c)^4$, where we have used the fact that $F(y_{t-d}; 0, c) = 0$.

[6] The tests are however easily generalized to the case where d is unknown by replacing y_{t-d} in (2) and (3) with a linear combination $\alpha' y_t^{(p)}$.

is nonzero only if the model is an LSTAR model[7], (ii) β_2 is zero if the model is an LSTAR model with $\theta_0 = c = 0$ but is always nonzero if the model is an ESTAR model, and (iii) that β_1 is zero if the model is an ESTAR model with $\theta_0 = c = 0$ but is always nonzero if the model is an LSTAR model. Combining these three properties of the auxiliary parameters leads to the following decision rule: if the p-value corresponding to H_{02} is the smallest, an ESTAR model should be selected, while in all other cases an LSTAR model is to be the preferred choice. The model selection procedure of Teräsvirta (1994) will be abbreviated as TP in the following.

Escribano and Jordá (1997) propose an alternative procedure which makes use of NL4 as test for general STAR type nonlinearity. Their decision rule to choose between the LSTAR and ESTAR alternatives is based on the observation that, assuming $\theta_0 = c = 0$ in (1), the properties of β_1 and β_2 given above also apply to β_3 and β_4 in (5), respectively. Therefore they suggest to test the hypotheses

$$H_{0E}: \qquad \beta_2 = \beta_4 = 0,$$

$$H_{0L}: \qquad \beta_1 = \beta_3 = 0,$$

and to select an LSTAR (ESTAR) model if the minimum p-value is obtained for H_{0L} (H_{0E}). This decision rule will be denoted EJP in the following.

Escribano and Jordá (1997,1998) present extensive simulation evidence on the relative performance of the linearity tests NL3 and NL4 and the decision rules TP and EJP. Their main findings can be summarized as follows. In case the true data generating process [DGP] is an LSTAR model, the power of the NL3 test in general is higher than the power of NL4, while the reverse holds if the DGP is an ESTAR model. This makes sense intuitively, as the p additional auxiliary regressors $\beta_4' \tilde{y}_t^{(p)} y_{t-d}^4$ in (5) are redundant in case of an LSTAR model, and the use of p extra degrees of freedom by NL4 causes a loss in power. In case of an ESTAR model, these extra terms contain vital information which more than compensates the use of additional degrees of freedom. Concerning the two decision rules to choose between LSTAR and ESTAR models, in general the performance of EJP appears superior over TP.

[7]Note that this is only true under the assumption that a first-order Taylor expansion is sufficient for the exponential function.

2.3 Robust tests and specification procedures

Van Dijk, Franses and Lucas (1998) [VDFL hereafter] analyze the properties of the linearity tests of LST in the presence of outliers. It is shown that in the case of a linear DGP with some additive outliers [AO's] the tests for STAR nonlinearity tend to reject the correct null hypothesis of linearity too often, even asymptotically. VDFL suggest to use outlier-robust estimation techniques using Generalized M [GM] estimators to estimate the model under the null hypothesis as a solution to this problem. As shown by VDFL, the resulting test statistics behave much better in the presence of AO's. For technical details we refer to VDFL, here we discuss only the intuition behind the GM estimation technique.

The GM estimator can be interpreted as an iterative weighted least squares procedure, where the weights are not fixed a priori but determined endogenously, in such a way that outliers are downweighted and do not influence the estimates of the parameters in the model under the null hypothesis. In addition to rendering better estimates of the null model, it allows to construct test statistics which are robust to outliers. Moreover, the weights assigned to the observations in the GM procedure can be used to detect aberrant data points. Examples of such use of these weights are provided in Section 5.

It is straightforward to robustify the TP and EJP decision rules in a similar manner, by simply using the GM estimator to estimate the models under the various null hypotheses (i.e. H_{03}, H_{02} and H_{01} in TP and H_{0L} and H_{0E} in EJP). The next two sections are devoted to investigating the properties of both the standard and robust linearity tests and specification procedures in the presence of outliers. In particular, we focus on whether the results from Escribano and Jordá (1997,1998) on the relative performance of the decision rules continue to hold in this case. We conjecture that whereas EJP might perform better than TP in case of no outliers, it might also be more sensitive to the presence of aberrant observations because of the inclusion of the term involving the fourth power of y_{t-d}.

3 Size of linearity tests and decision rules in the presence of outliers

This section adresses the first question raised in the Introduction by examining the behavior of the regular linearity tests NL3 and NL4 and decision rules TP and EJP in case of a linear data generating process [DGP] with AO's. In particular, the following DGP is used,

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t, \tag{6}$$

$$y_t = x_t + \xi\delta_t, \quad t = 1, \ldots, T, \tag{7}$$

with $\varepsilon_t \sim$ i.i.d.$N(0, \sigma^2)$, ξ a positive constant, and $P(\delta_t = 1) = P(\delta_t = -1) = \pi/2$, $P(\delta_t = 0) = 1 - \pi$, for some $0 \leq \pi < 1$. Hence, the observed series y_t consists of the core process x_t, which is generated according to an autoregressive (AR) process of order two[8], and a contamination process $\xi\delta_t$.

We consider three parameterizations for the core process x_t, by varying $\phi = (\phi_0, \phi_1, \phi_2)'$ among $\phi = (0, .2, .2)'$, $\phi = (0, .65, .2)'$ and $\phi = (0, 1.4, -.65)'$, denoted as DGP I, II, and III, respectively. The residual variance σ^2 is set equal to unity throughout. The AR coefficients are chosen such that they are in different parts of the stationary region of the parameter space for (ϕ_1, ϕ_2) for the different DGP's. The roots of the AR polynomial are real and relatively small (DGP I), real and large (DGP II) and complex and large (DGP III). Besides the case where $\pi = 0$ (no outlier contamination), we set $\xi = 3\sigma, 5\sigma, 7\sigma$ and $\pi = .01, .05$, rendering 7 experiments per DGP. We use 5000 replications for each experiment and take $T = 300$ as the sample size. Necessary starting values are always set equal to zero, while the first 100 observations in the artificial samples are discarded in order to eliminate any possible influence of this choice. Finally, in all experiments, the AR-order is assumed known[9].

- insert Table 1 about here -

Table 1 shows the rejection frequencies of the NL3 and NL4 tests with y_{t-1} as candidate transition variable at a significance level of 5%, using asymptotic critical values. It is seen that in case no outliers are present, both tests are somewhat undersized, which corroborates the findings of LST, among others. As soon as outliers are added, the actual rejection frequencies of the tests exceed the nominal significance level of 5%. For the values of ξ and π considered here it is seen that the magnitude of the size distortion increases as either the

[8] We focus on AR(2) (and STAR(2) models in Section 4) as the first-order case is covered extensively in VDFL.

[9] Obviously, the presence of outliers might affect the behavior of commonly used order selection criteria such as the Akaike and Schwarz Information Criteria as well, see Ronchetti (1997) for a recent overview. This point however is besides the objective of this paper.

magnitude or frequency of outliers increases. It should be remarked that VDFL show that as $\xi \to \infty$, the asymptotic distribution of the test statistics returns to a χ^2 distribution again, i.e., for very large values of ξ the size distortion disappears. However, in that case the power of the tests also collapses to their size.

- insert Figure 1 about here -

To demonstrate that our findings are not specific for the 5% significance level which is used, Figure 1 shows p-value discrepancy plots for NL3 for DGP's I and II. These plots, advocated by Davidson and MacKinnon (1998), graph the difference between the actual and nominal size of the tests versus the nominal size[10]. It is seen that the plots lie completely above the zero line for all values of π and ξ, i.e., the actual size is always larger than the nominal size. Finally, comparing the rejection frequencies of NL3 and NL4 seems to suggest that the size distortions are of comparable magnitude.

- insert Table 2 about here -

Table 2 displays the frequency of selecting an ESTAR model by TP and EJP, conditional upon rejecting linearity by NL3 and NL4, respectively. First of all, note that TP seems to be biased toward selecting LSTAR models in case of DGP's I and III, as the frequency of selecting an ESTAR alternative is well below 1/2 in case no outliers are present. When the series are contaminated with AO's, both decision rules become biased toward selecting the ESTAR alternative, which becomes more and more evident when the magnitude or the frequency of occurrence of outliers increases. This might be explained by the symmetric nature of the contamination process, as this creates both very small and very large aberrant observations which might mimic ESTAR data.

[10]To be more precise, these p value discrepancy plots are constructed as follows. The N replications in the Monte Carlo experiments render p-values $p_1, \ldots, p_N$, where in our case $N = 5000$. The empirical distribution function of the p-values can be estimated by simply calculating $\hat{F}(x) = \frac{1}{N}\sum_{j=1}^{N} I(p_j \leq x)$, for any point x in the (0,1) interval, where $I(A)$ denotes the indicator function for the event A. The function $\hat{F}(x)$ gives the actual rejection frequency of the test at nominal significance level x. If the distribution used to calculate the p-values p_j is correct, each of the the p_j should be distributed as uniform (0,1), and $\hat{F}(x) \approx x$. By calculating $\hat{F}$ on a grid of points $x_1, \ldots, x_M$ on the (0,1) interval and plotting $\hat{F}(x_i) - x_i$ against x_i one can easily infer if the test statistic is under- or oversized at various different nominal significance levels. Moreover, it allows easy comparison between different test statistics.

4 Power of linearity tests and decision rules

This section discusses the behavior of the linearity tests NL3 and NL4 and the TP and EJP decision rules in case the DGP is a STAR model, possibly contaminated with outliers. For this purpose, we replace the AR(2) model for the core process x_t in (6) by a STAR(2) model

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + F(x_{t-1}; \gamma, c)(\theta_0 + \theta_1 x_{t-1} + \theta_2 x_{t-2}) + \varepsilon_t, \tag{8}$$

where $F(x_{t-1}; \gamma, c)$ is either the logistic or exponential function (where it should be noted that we do not subtract 1/2 from the logistic function here) and $\varepsilon_t \sim$ i.i.d.$N(0, \sigma^2)$. We first consider the model used by Teräsvirta, Lin and Granger (1993) and Teräsvirta (1994) where $\phi_0 = 0$, $\phi_1 = 1.8$, $\phi_2 = -1.06$, $\theta_1 = -.9$, $\theta_2 = .795$, $\gamma = 100$ or 1000 in the LSTAR and ESTAR case, respectively, and $\sigma = .02$.

Notice that the roots of the AR model which results when F is equal to 0 in (8) are complex and larger than unity in modulus. Hence, the STAR model is explosive in this regime, and the series x_t has the tendency to be propelled back to more stable parts of the state space. The properties of the time series generated by the model now crucially depend on the relative magnitude of θ_0 and c, as they jointly determine the value of the attractor x^* of the model[11] and its (in)stability. This also implies that the model can generate very 'asymmetric' realizations in the sense that the distribution of the observations over the different regimes can be very asymmetric. In fact, the data may seem to have been generated by linear outlier-type models. For example, if $c = \theta_0 = 0$, the attractor of (8) is equal to $x^* = 0$. In the LSTAR model $F(x^*; \gamma, c) = 1/2$, and the attractor is stable since the roots of the effective AR(2) polynomial at the attractor are smaller than unity. In the ESTAR model on the other hand $F(x^*; \gamma, c) = 0$, the attractor is unstable, and the model generates 'endogenous fluctuations', i.e., if the residual process is 'switched off', the resulting time series does not converge to the attractor but keeps on fluctuating around zero in a limit cycle. As another example, if $c = .02$ and $\theta = .04$ (which is one of the cases considered in Teräsvirta, 1994), the LSTAR model has a stable attractor at $x^* = 0.10$ for which the transition function takes the value 1. In fact, observations for which $x_t < c$ are rare since the AR model is explosive in that region, and realizations

[11] The attractor(s) of a model $x_t = f(x_{t-1}, \ldots, x_{t-p}) + \varepsilon_t$ is (are) points x^* for which $x^* = f(x^*, \ldots, x^*)$.

from this model are very asymmetric. The same asymmetry is observed in the ESTAR model. Although this model now has multiple fixed points, only the largest of those is stable, and the majority of observations is centered around this point. Observations for which the value of the transition function is smaller than .5 are rare. Concluding, the model is capable of generating series which closely resemble linear time series with outliers. We conjecture that it should be very difficult to distinguish between the two in such cases. In our experiments we consider the following combinations of θ_0 and c: $(\theta_0, c) = (0,0), (.02,0), (0,.02), (.02,.02)$, and $(.04,.02)$. Again we set the sample size equal to $T = 300$, while the magnitude and frequency of occurrence of outliers are taken as $\xi = 3\sigma, 5\sigma$ and 7σ, and $\pi = .05$. For each experiment, 1000 replications are used[12].

- insert Table 3 about here -

Table 3 shows the actual rejection frequencies of both the standard tests and their robust counterparts at the 5% nominal significance level. First of all, it is seen that the power of both the standard and robust tests is excellent in case no outliers are present, except for the LSTAR model with $\theta_0 = .04$ and $c = .02$. Evidently this is caused by the highly asymmetric properties of series generated by this model as discussed above. Note that this is also the single case in which the use of the robust tests causes a substantial loss in power. When outliers are added, the power of the tests remains satisfactory, although in some cases the power of the robust tests is reduced quite considerably. Adding outliers to the LSTAR model with $\theta_0 = .04$ and $c = .02$ increases the power of the standard tests dramatically, whereas the power of the robust tests is affected to a much lesser extent.

- insert Table 4 about here -

The frequencies of selecting the correct model by the two decision rules, conditional upon rejecting linearity by the respective general linearity tests, are set out in Table 4. The presence of outliers causes the performance of the standard decision rules to deteriorate in case the DGP is an LSTAR model, in the sense that the correct model is chosen less and less often when the magnitude of outliers increases[13]. The effect on the robust procedures

[12] Results for other sample sizes and frequencies of occurrence of outliers are available on request from the corresponding author.

[13] Additional simulations with different values of π (not shown here) suggest that the same effect occurs when the frequency of occurrence of outliers increases.

is much less pronounced. However, we arrive at the opposite conclusion when the DGP is an ESTAR model: now the decision rules based on the regular LM tests continue to behave well while the performance of the robust procedures deteriorates.

- insert Tables 5 and 6 about here -

Next we consider a parameterization of (8) for which the resulting AR(2) model is stable for all values of the transition function. In particular, we set $\phi_0 = 0$, $\phi_1 = 1.4$, $\phi_2 = -.65$, $\theta_1 = -.9$, $\theta_2 = .79$, $\gamma = 10$ and $\sigma = .2$. Because the model is stable for all possible values of the transition function, θ_0 and c do not exercise such a large influence on the properties of time series generated by the model. For that reason we only consider the combinations $(\theta_0, c) = (0, 0)$, $(.02, 0)$, and $(0, .02)$. Table 5 shows the actual rejection frequencies of both the standard tests and their robust counterparts at 5% nominal significance level, while Table 6 displays the frequencies of selecting the correct model, conditional upon rejecting linearity. Table 5 demonstrates that both the NL3 and NL4 tests work well in this case, although power diminishes somewhat in case outliers are added. From Table 6, we observe that again the standard specification procedures break down in case an LSTAR model is the true DGP. The robust specification procedures on the other hand keep selecting the correct model in the majority of cases. The opposite also holds here in case the DGP is an ESTAR model. Some guidelines of how to proceed in practice are given in Section 6.

5 Empirical illustrations

In this section we demonstrate the effects of outliers on the specification procedures for STAR models by applying them to monthly gold and silver prices and to several real exchange rate series.

5.1 Gold and silver prices

The gold and silver price data are taken from the IMF *International Financial Statistics* and cover the period from November 1971 (when the prices of these precious metals were deregulated) until June 1994 (272 observations). The price of gold is US$ per fine ounce in London, while the price of silver is US$ per troy ounce in New York. The same data set has been analyzed in Escribano and Granger (1998), who focus on the possible existence of a long-run relationship between these prices and make use of cointegration techniques, allowing for possible nonlinear error-correction.

The raw gold and silver price series are displayed in the upper panels of Figures 2 and 3, respectively. Some remarkable features of these series clearly stand out from these graphs. In particular, the price of silver sharply increases in the second half of 1979 and returns to previous levels some six to nine months later. The price of gold increases at about the same time, but appears to remain at a higher level afterwards. The sudden and large increase in the price of silver can be explained by the speculative attack on the silver market by the Hunt brothers from Texas (and others). Although the attack was not aimed at the gold market directly, the increase in the gold price at the time of the 'bubble' in the silver price seems to suggest that it did have an effect on the gold price. Although this might suggest a possible connection between the two markets, Escribano and Granger (1998) report serious difficulties in extracting the particular form of the relationship, and suggest that the markets may have become increasingly separated, especially since 1990. Here we completely ignore the issue of a possible linkage between the two markets and restrict ourselves to univariate analysis of both time series.

We perform our analysis on the associated return series, which are constructed by taking first differences of logarithms of the respective prices. First, we fit univariate AR models to both returns on silver and gold. For both return series an AR(2) model is suggested by the Schwarz Information Criterion [SIC], and this seems sufficient to capture the correlation properties of the series. The estimates of the parameters in the AR(2) models are not relevant for our purpose and are therefore not shown here. We proceed by testing for linearity by subjecting the series to the linearity tests NL3 and NL4 for values of $d = 1, \ldots, 6$. The p-values for the different tests are given in Tables 7 and 8 for the gold and silver returns respectively. These tables also report the p-values corresponding to the different tests which are carried out in the TP and EJP decision rules.

- insert Tables 7 and 8 about here -

It is seen that the standard tests decisively reject the null hypothesis of linearity, for different values of the delay parameter d. The outcomes of the test sequences give little guidance as to which model is most appropriate for either the gold or silver return series; not only do the two decision rules contradict each other in quite a few cases, the preferred model also crucially depends on the value of d which is selected. This might tentatively be interpreted as evidence that something else than genuine STAR nonlinearity is going on. This intuition is corroborated by the finding that we have tried to estimate both LSTAR and ESTAR models for several different choices of d, but have not been able to obtain any sensible results.

The lower halves of Tables 7 and 8 show results from applying the robust tests for STAR nonlinearity and the robust specification procedures. It is seen that the conclusion is radically different from the standard procedures: linearity can hardly be rejected at conventional significance levels, only for $d = 3$ for the silver returns and $d = 1$ for the gold returns is there some indication for the presence of nonlinearity. Also note that in both of these cases EJP is not very informative about the preferred type of model.

- insert Figures 2 and 3 about here -

The middle panel of Figures 2 and 3 show the series of gold and silver returns, respectively. We have marked the observations which are downweighted in the robust estimation procedure for the linear AR(2) model which is assumed to hold under the null hypotheses corresponding to NL3 and NL4. The lower panels display the actual weights. It is seen that for the silver return series, all observations during and surrounding the speculative attack in 1979-1980 receive a weight equal to zero, i.e., they are regarded as obvious outliers and their influence on the parameter estimates is eliminated completely. In addition, the middle of 1982 and the beginning of 1983 also appear to be periods of relative unrest and aberrant observations. For the gold return series, the 1979-1980 period does not stand out so clearly, although the estimation procedure signals that it contains some outlying observations. Additionally, 1973-1974 is a period in which quite some outliers have occurred. Possibly this is related to the end of the Bretton Woods era and/or the first oil crisis.

Of course one might have suggested beforehand that the time series over the sample period considered here contain some unusual events and probably are contaminated with some aberrant observations and, hence, that caution should be exercised before interpreting any statistical test results. The main advantage of the robust procedures is that it automatically guards against the presence of malignant observations, without requiring a priori knowledge by the researcher concerning their timing.

5.2 Real exchange rates

The question whether or not purchasing power parity [PPP] holds as a long run equilibrium relationship has been heavily debated in recent years, see Froot and Rogoff (1995) and Rogoff (1996) for surveys. One of the reasons for this controversy is that in general standard unit root tests fail to reject nonstationarity of real exchange rates when applied to data from the post-Bretton Woods era. Several statistical explanations have been put forward for this failure to find evidence in favor of PPP, such as the lack of power of standard unit root tests in small samples and the lack of power against near-nonstationary alternatives. The proposed solutions include the use of long spans of data (Grilli and Kaminsky 1991, Lothian and Taylor 1996) and the use of panel unit root tests (Frankel and Rose 1996, O'Connell 1998). Economic explanations have also been given, such as the presence of non-traded goods in the price indices which are usually employed to construct real exchange rates (Rogers and Jenkins 1995) and the presence of transaction costs (Davutyan and Pippenger 1990, Dumas 1992, Uppal 1993). Here we focus on the latter.

The presence of transaction costs leads to the notion of different regimes in real exchange rates. In particular, the profits from commodity arbitrage, which is generally thought to be the ultimate force behind maintaining PPP, do not make up for the costs involved in the necessary transactions for small deviations from the equilibrium real exchange rate. This implies the existence of a band around the equilibrium rate in which there is no tendency of the real exchange rate to revert to its equilibrium value. Outside this band, commodity arbitrage becomes profitable, which forces the real exchange rate back towards the band. For an analytic derivation of an equilibrium model of exchange rate determination which takes these effects of transaction costs into account we refer to Dumas (1992).

Several approaches have been employed to examine the importance of this transaction costs argument. For example, Rogers and Jenkins (1995), Engel and Rogers (1996), and Jenkins (1997), among others, approximate transaction costs with variables such as distance between countries or cities. Obstfeld and Taylor (1997) and O'Connell and Wei (1997) apply the BAND-TAR model of Balke and Fomby (1997) to model the no-reversion band around the equilibrium real exchange rate explicitly. Michael *et al.* (1997) argue that (E)STAR models might also be capable to describe the dynamic properties of real exchange rates in the presence of transaction costs. Using a two-century span of annual data and a sample of monthly interwar exchange rates they find moderate evidence in favor of their maintained hypothesis. Baum *et al.* (1998) apply the same methodology to real exchange rates from the post-Bretton Woods era, and find "clear empirical support for the presence of nonlinearities in the dynamic adjustment to deviations from PPP" (p. 15). Here we apply the standard and robust LM-type linearity tests and decision rules to check whether their findings may be due to the presence of outliers.

The data are taken from the IMF *International Financial Statistics* and cover the period from January 1973 until December 1994 (264 monthly observations). We consider exchange rates of a large number of industrialized countries vis-à-vis the US dollar and use Consumer Price Indices [CPI] to construct the real exchange rates[14]. Furthermore, we impose homogeneity conditions, i.e., the (log) real exchange rate q_t is constructed as $q_t = s_t - p_t^* + p_t$, where s_t is the log of the nominal exchange rate (units of foreign currency per US dollar) and p_t^* and p_t are logs of the foreign and domestic (US) wholesale price indices.

We calculate the test statistics after transforming the STAR model into the familiar Dickey-Fuller format, i.e., we test for linearity of

$$\Delta q_t = \phi_0 + \phi_p q_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta q_{t-i} + \left(\theta_0 + \theta_p q_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta q_{t-i} \right) \times (F(q_{t-d}; \gamma, c) + \varepsilon_t. \quad (9)$$

The AR-order p is determined using the 'general-to-specific' approach advocated by Ng and Perron (1995) in the context of unit root tests, see also Hall (1994). We start with an initial number of lagged first differences $p_{\max}$ and then sequentially test, using conventional t-statistics at a prespecified significance level α, for the statistical significance of the highest order lag, using a backward elimination algorithm. The order p is then selected as the maximum lag length at which the algorithm terminates. The reason for not relying on information criteria in this case is that the orders which are suggested by, for example, the Schwarz Information Criterion [SIC] are far too small to adequately capture the serial correlation properties of the series.

Table 9 reports the miminum p-values for the linearity tests NL3 and NL4 as well as their robust variants, where the minimum is taken over $d = 1, \ldots, 6$. The selected AR-orders, which are determined by the Ng and Perron (1995) procedure with $p_{\max} = 12$

[14] Baum *et al.* (1998) also consider real exchange rates based on wholesale price indices. Results from the regular and robust specification procedures for these series are qualitatively similar to the results presented here for the CPI-based measures and are available upon request from the corresponding author.

and $\alpha = .90$, are given as well. A general conclusion which emerges from this table is that the p-values for the standard tests are in general lower than the p-values of their robust counterparts, often very much so. Using a significance level of 10%, for 7 (11) of the countries considered, the NL3 (NL4) rejects the null hypothesis of linearity, while the RNL3 (RNL4) does not. Table 10 displays the models which are selected by the various decision rules, where 'Linear' is reported if the p-value of the corresponding general linearity test is larger than 0.10. The selected delay parameters d are reported in parentheses. It is seen that especially the Escribano-Jordá rule tends to select an LSTAR model, whereas an ESTAR model would be the most obvious choice in light of the transactions costs argument. This may be due of course to the fact that LSTAR and ESTAR models can be close substitutes, especially if in case of an ESTAR model the majority of the observations in the regime where $F \approx 1$ lies in one of the regions $y_{t-d} >> c$ and $y_{t-d} << -c$.

- insert Tables 9 and 10 about here -

If we focus on the robust tests, the general conclusion from Tables 9 and 10 is that real exchange rates appear to be linear, and that these series once in a while show substantial outliers.

To substantiate this conclusion, Figures 4 to 6 provide some information about the results of estimating the linear model under the null hypothesis by the GM procedure for the Finnish markka, Norwegian kroner, and Swedish kroner real exchange rates. The middle panels of these Figures graph the first differences of the log real exchange rates, where observations which are downweighted are marked with circles. The lower panels again display the actual weights.

- insert Figures 4-6 about here -

It is clear from these Figures (which are representative for the all real exchange rate series) that only very few observations are downweighted, especially surrounding the second oil crises in 1979, the large increase in the exchange rate of the US dollar around 1985, and the turbulence in the EMS exchange rate system in 1991-1992. In fact, closer inspection of the weights for the various series reveals that they contain quite some 'common outliers'. To provide some information on those common outliers, consider Figure 7, which graphs the series N_t, defined as the number of countries for which the observation at time t receives a weight less than one.

- insert Figure 7 about here -

The observations in July 1975, October 1978, November 1978, March 1980, July 1985, March 1991, December 1991, and October 1992 are downweighted for half or more of the real exchange rates[15].

Concluding, the suggestion of nonlinear adjustment in the real exchange rate is caused by only a small number of data points. The fact that a large part of these data points coincide for the different series strengthens the conclusion that they do not signal intrinsic nonlinearity in the process generating the real exchange rates, but rather are caused by some aberrant exogenous events. Hence, the presence of transaction costs does not seem to imply nonlinear behavior of real exchange rates.

6 Summary and concluding remarks

This paper has compared the relative performance of the specification procedures for STAR models proposed by Teräsvirta (1994) and Escribano and Jordá (1997), as well as outlier-robust variants. Various circumstances have been considered, i.e., linear and nonlinear DGP's, with and without outliers. This final section aims to provide some practical guidelines how to proceed in practice, when one cannot be sure whether certain features of a particular time series under scrutiny are caused by genuine nonlinearity or by some outliers. It is suggested that both standard and robust linearity tests and specification procedures are applied, and that the outcomes are compared to reach a conclusion. The Monte Carlo evidence presented in Sections 3 and 4 suggests the following 'decision rules'. If both the standard and robust tests do not reject the null hypothesis of linearity, one can be reasonably confident that the DGP of the series is linear. When both standard and robust tests reject the null hypothesis, one might assume that the DGP of the series is genuinely nonlinear - although it is possible of course that it is linear with a high frequency of occurrence of large outliers, such that the sizes of both test procedures are heavily distorted. The case where the standard tests reject linearity while the robust tests do not points towards the possibility that the nonlinearity which is detected by the standard test procedures is caused by only a few outliers. A further investigation of the series, especially the 'influential' observations (i.e., those which are downweighted by the robust

[15]The exact numbers are 13, 11, 15, 9, 12, 16, 10, and 13, respectively.

estimation procedure) is strongly called for. The fact that the robust estimation procedure endogenously determines the weights for the different observations is seen to be advantageous once again here, as this allows one to easily determine which observations cause the standard tests to reject the null hypothesis. Alternatively, one might have encountered a case where the DGP is nonlinear but contaminated in such a way that the ‘power’ of the standard test increases while the power of the robust test does not. Also in this case it is advisable to further investigate the series for the presence of outliers before estimating a nonlinear model. Finally, if the standard test does not reject the null, while the robust test does, it is perhaps most likely that the DGP is nonlinear with some contamination such that the power of the standard test is decreased.

References

Anderson, H.M., 1997, Transaction costs and nonlinear adjustment towards equilibrium in the US Treasury Bill market, *Oxford Bulletin of Economics and Statistics* **59**, 465–484.

Balke, N.S. and T.B. Fomby, 1997, Threshold cointegration, *International Economic Review* **38**, 627–646.

Baum, C.F., M. Caglayan and J.T. Barkoulas, 1998, Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era, Working paper No. 404, Department of Economics, Boston College.

Davidson, R. and J.G. MacKinnon, 1998, Graphical methods for investigating the size and power of hypothesis tests, *The Manchester Scool* **66**, 1–26.

Davutyan, N. and J. Pippenger, 1990, Testing purchasing power parity: some evidence of the effects of transaction costs, *Econometric Reviews* **9**, 211–240.

Dumas, B., 1992, Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* **5**, 153–180.

Engel, C. and J.H. Rogers, 1996, Regional patterns in the law of one price: the roles of geography vs. currencies, International Finance Discussion Papers No. 533, Federal Reserve.

Escribano, A. and C.W.J. Granger, 1998, Investigating the relationship between gold and silver prices, *Journal of Forecasting* **17**, 81–107.

Escribano, A. and O. Jordá, 1997, Testing nonlinearity: decision rules for selecting between logistic and exponential STAR models, Working paper, Universidad Carlos III de Madrid.

Escribano, A. and O. Jordá, 1998, Improved testing and specification of smooth transition regression models, in P. Rothman (editor), *Nonlinear Time Series Analysis of Economic and Financial Data*, Boston: Kluwer Academic Press, forthcoming.

Frankel, J.A. and A.K. Rose, 1996, A panel prject on purchasing power parity: mean reversion within and between countries, *Journal of International Economics* **40**, 209–224.

Froot, K.A. and K. Rogoff, 1995, Perspectives on PPP and long-run real exchange rates, in G. Grossman and K. Rogoff (editors), *Handbook of International Economics, vol. 3*, Amsterdam: North-Holland, pp. 1647–1688.

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

Grilli, V. and G. Kaminsky, 1991, Nominal exchange rate regimes and the real exchange rate, *Journal of Monetary Economics* **27**, 191–212.

Hall, A., 1994, Testing for a unit root in time series with pretest data-based model selection, *Journal of Business and Economic Statistics* **12**, 461–470.

Jansen, E.S. and T. Teräsvirta, 1996, Testing parameter constancy and super exogeneity in econometric equations, *Oxford Bulletin of Economics and Statistics* **58**, 735–768.

Jenkins, M.A., 1997, Cities, borders, distances, non-traded goods and purchasing power parity, *Oxford Bulletin of Economics and Statistics* **59**, 203–213.

Lothian, J. and M.P. Taylor, 1996, Real exchange rate behavior: the recent float from the perspective of the past two centuries, *Journal of Political Economy* **104**, 488–509.

Luukkonen, R., P. Saikkonen and T. Teräsvirta, 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Michael, P., A.R. Nobay and D.A. Peel, 1997, Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation, *Journal of Political Economy* **105**, 862–879.

Ng, S. and P. Perron, 1995, Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association* **90**, 268–281.

Obstfeld, M. and A.M. Taylor, 1997, Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited, NBER Working paper No. 6053.

O'Connell, P.G.J., 1998, The overvaluation of purchasing power parity, *Journal of International Economics* **44**, 1–19.

O'Connell, P.G.J. and S.-J. Wei, 1997, "The bigger they are, the harder they fall"; how price differences across U.S. cities are arbitraged, NBER working paper No. 6089.

Rogers, J.H. and M. Jenkins, 1995, Haircuts or hysteresis? Sources of movements in real exchange rates, *Journal of International Economics* **38**, 339–360.

Rogoff, K., 1996, The purchasing power parity puzzle, *Journal of Economic Literature* **34**, 647–668.

Ronchetti, E., 1997, Robustness aspects of model choice, *Statistica Sinica* **7**, 327–338.

Skalin, J. and T. Teräsvirta, 1996, Another look at Swedish business cycles, Working Paper Series in Economics and Finance No. 130, Stockholm School of Economics.

Skalin, J. and T. Teräsvirta, 1998, Modelling asymmetries in unemployment rates, *mimeo*, Stockholm School of Economics.

Teräsvirta, T., 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

Teräsvirta, T., 1998, Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (editors), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

Teräsvirta, T., C-F.J. Lin and C.W.J. Granger, 1993, Power of the neural network linearity test, *Journal of Time Series Analysis* **14**, 209–220.

Teräsvirta, T. and H.M. Anderson, 1992, Characterizing nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics* **7**, S119–S136.

Uppal, R., 1993, A general equilibrium model of international portfolio choice, *Journal of Finance* **48**, 529–553.

Van Dijk, D. and P.H. Franses, 1998a, Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics*, forthcoming.

Van Dijk, D. and P.H. Franses, 1998b, Nonlinear error-correction models for interest rates in the Netherlands, in S. Hylleberg (editor), *Nonlinear dynamics*, Cambridge: Cambridge University Press, forthcoming.

Van Dijk, D., P.H. Franses and A. Lucas, 1998, Testing for smooth transition nonlinearity in the presence of additive outliers, *Journal of Business and Economic Statistics*, forthcoming.

Wolters, J., T. Teräsvirta and H. Lütkepohl, 1996, Modelling the demand for M3 in the unified Germany, Working Paper Series in Economics and Finance No. 113, Stockholm School of Economics.

Table 1: Size of nonlinearity tests in the presence of additive outliers[1]

		DGP I		DGP II		DGP III	
π	ζ	NL3	NL4	NL3	NL4	NL3	NL4
.00	0	0.041	0.040	0.042	0.046	0.044	0.050
.01	3	0.062	0.068	0.173	0.167	0.282	0.280
	5	0.124	0.120	0.705	0.676	0.779	0.782
	7	0.168	0.158	0.920	0.911	0.932	0.941
.05	3	0.096	0.095	0.411	0.394	0.455	0.441
	5	0.161	0.186	0.939	0.965	0.958	0.959
	7	0.172	0.197	0.986	0.998	0.996	0.999

[1] Size of nonlinearity tests NL3, NL4. Series are generated according to (6) and (7). The parametrizations for the core processes in DGP I-III are given by $\phi = (\phi_0, \phi_1, \phi_2)' = (0, .2, .2)'$, $\phi = (0, .65, .2)'$ and $\phi = (0, 1.4, -0.65)'$ with $\sigma = 1$, respectively. The table is based on 5000 replications for sample size $T = 300$.

Table 2: Behavior of decision rules in the presence of additive outliers[1]

		DGP I		DGP II		DGP III	
π	ζ	TP	EJP	TP	EJP	TP	EJP
.00	0	0.396	0.530	0.502	0.570	0.384	0.452
.01	3	0.422	0.529	0.593	0.591	0.502	0.536
	5	0.460	0.605	0.685	0.730	0.603	0.673
	7	0.442	0.612	0.647	0.777	0.641	0.788
.05	3	0.530	0.687	0.771	0.780	0.703	0.729
	5	0.583	0.805	0.899	0.954	0.900	0.956
	7	0.587	0.792	0.925	0.986	0.938	0.995

[1] Frequency of selecting an ESTAR model by the decision rules of Teräsvirta (1994) [TP] and Escribano and Jordá (1997) [EJP], conditional upon rejecting linearity by NL3 and NL4, respectively. Series are generated according to (6) and (7). The parametrizations of the core process for DGP I-III are given by $\phi = (\phi_0, \phi_1, \phi_2)' = (0, .2, .2)'$, $\phi = (0, .65, .2)'$ and $\phi = (0, 1.4, -0.65)'$ with $\sigma = 1$, respectively. The table is based on 5000 replications for sample size $T = 300$.

Table 3: Power of nonlinearity tests in the presence of additive outliers[1]

				LSTAR				ESTAR			
θ_0	c	π	ξ	NL3	RNL3	NL4	RNL4	NL3	RNL3	NL4	RNL4
0.00	0.00	.00	0	1.000	1.000	1.000	1.000	0.999	0.997	0.999	0.998
		.05	3	0.972	0.978	0.967	0.981	0.972	0.861	0.966	0.935
			5	0.951	0.968	0.963	0.968	0.998	0.869	0.999	0.939
			7	0.987	0.991	0.997	0.991	0.999	0.971	0.998	0.997
0.02	0.00	.00	0	0.989	0.952	0.988	0.946	1.000	1.000	1.000	1.000
		.05	3	0.904	0.813	0.904	0.803	0.984	0.941	0.979	0.969
			5	0.966	0.855	0.966	0.828	0.998	0.960	0.999	0.973
			7	0.995	0.935	0.997	0.946	0.999	0.996	1.000	1.000
0.00	0.02	.00	0	1.000	1.000	1.000	1.000	0.993	0.985	0.996	0.995
		.05	3	0.992	0.994	0.992	0.995	0.913	0.903	0.918	0.930
			5	0.954	0.984	0.964	0.988	0.992	0.859	0.997	0.918
			7	0.988	0.987	0.991	0.989	0.997	0.747	1.000	0.848
0.02	0.02	.00	0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		.05	3	0.981	0.980	0.973	0.970	0.976	0.970	0.985	0.985
			5	0.938	0.969	0.934	0.971	0.997	0.957	1.000	0.979
			7	0.986	0.996	0.986	0.997	0.999	0.930	1.000	0.965
0.04	0.02	.00	0	0.403	0.220	0.396	0.207	1.000	1.000	1.000	1.000
		.05	3	0.751	0.567	0.730	0.548	0.998	0.993	0.995	0.997
			5	0.980	0.515	0.994	0.494	0.999	0.992	0.998	0.996
			7	0.985	0.286	0.999	0.274	0.999	1.000	0.999	1.000

[1] Power of nonlinearity tests NL3, NL4 and their outlier-robust counterparts, RNL3 and RNL4. Series are generated according to (8) and (7) with $\phi_0 = 0$, $\phi_1 = 1.8$, $\phi_2 = -1.06$, $\theta_1 = -0.9$, $\theta_2 = 0.795$, $F(y_{t-d}; \gamma, c)$ equal to the logistic function (2) with $d = 1$, $\gamma = 100$ [LSTAR], or equal to the exponential function (3) with $d = 1$, $\gamma = 1000$, and $\sigma = 0.02$. The table is based on 1000 replications of length $T = 300$.

Table 4: Behavior of decision rules in the presence of additive outliers[1]

				LSTAR				ESTAR			
θ_0	c	π	ξ	TP	RTP	EJP	REJP	TP	RTP	EJP	REJP
0.00	0.00	.00	0	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
		.05	3	0.829	0.911	0.990	0.993	0.952	0.865	0.989	0.991
			5	0.393	0.892	0.720	0.964	0.953	0.675	0.995	0.991
			7	0.156	0.967	0.270	0.978	0.933	0.885	0.998	1.000
0.02	0.00	.00	0	0.993	0.984	0.982	0.949	0.995	0.980	0.965	0.957
		.05	3	0.898	0.916	0.827	0.788	0.935	0.697	0.945	0.890
			5	0.549	0.847	0.644	0.842	0.947	0.461	0.988	0.894
			7	0.254	0.968	0.317	0.925	0.906	0.786	0.972	0.944
0.00	0.02	.00	0	0.999	0.997	1.000	1.000	0.861	0.453	0.477	0.109
		.05	3	0.806	0.973	0.986	0.943	0.875	0.436	0.828	0.305
			5	0.398	0.926	0.755	0.900	0.922	0.343	0.992	0.280
			7	0.153	0.931	0.333	0.899	0.943	0.252	0.999	0.225
0.02	0.02	.00	0	0.997	0.998	0.998	0.983	0.856	0.609	0.416	0.104
		.05	3	0.854	0.962	0.919	0.903	0.842	0.539	0.432	0.211
			5	0.547	0.955	0.571	0.811	0.927	0.556	0.639	0.279
			7	0.272	0.964	0.170	0.767	0.941	0.504	0.824	0.426
0.04	0.02	.00	0	0.953	0.795	0.674	0.734	0.841	0.711	0.057	0.059
		.05	3	0.530	0.342	0.616	0.768	0.896	0.469	0.552	0.413
			5	0.296	0.384	0.823	0.836	0.934	0.280	0.819	0.526
			7	0.198	0.780	0.933	0.708	0.816	0.470	0.481	0.192

[1] Relative frequencies of correctly choosing the correct type of model from the decision rules described in Section 2.2. Frequencies are conditional on rejecting the null hypothesis of linearity by the general test corresponding to the decision rules (e.g. the entries in the column EJP are conditional on rejection of linearity by test NL4). Series are generated according to (8) and (7) with $\phi_0 = 0$, $\phi_1 = 1.8$, $\phi_2 = -1.06$, $\theta_1 = -0.9$, $\theta_2 = 0.795$, $F(y_{t-d}; \gamma, c)$ equal to the logistic function (2) with $d = 1$, $\gamma = 100$ [LSTAR], or equal to the exponential function (3) with $d = 1$, $\gamma = 1000$, and $\sigma = 0.02$. The table is based on 1000 replications of length $T = 300$.

Table 5: Power of nonlinearity tests in the presence of additive outliers[1]

				LSTAR				ESTAR			
θ_0	c	π	ξ	NL3	RNL3	NL4	RNL4	NL3	RNL3	NL4	RNL4
0.00	0.00	.00	0	0.997	0.997	0.998	0.996	0.952	0.938	0.971	0.969
		.05	3	0.899	0.887	0.905	0.909	0.928	0.939	0.963	0.968
			5	0.981	0.922	0.994	0.930	0.976	0.845	0.998	0.925
			7	0.992	0.987	1.000	0.977	0.974	0.771	0.998	0.889
0.02	0.00	.00	0	0.997	0.998	0.996	0.996	0.952	0.941	0.975	0.969
		.05	3	0.886	0.882	0.892	0.892	0.924	0.935	0.963	0.970
			5	0.981	0.914	0.992	0.924	0.977	0.843	0.998	0.927
			7	0.991	0.981	1.000	0.976	0.975	0.784	0.998	0.906
0.00	0.02	.00	0	0.998	0.997	0.997	0.996	0.949	0.941	0.975	0.973
		.05	3	0.901	0.896	0.910	0.916	0.930	0.939	0.963	0.971
			5	0.984	0.929	0.994	0.926	0.976	0.848	0.997	0.928
			7	0.992	0.981	1.000	0.976	0.972	0.766	0.998	0.906

[1] Power of nonlinearity tests NL3, NL4 and their outlier-robust counterparts, RNL3 and RNL4. Series are generated according to (8) and (7) with $\phi_0 = 0$, $\phi_1 = 1.4$, $\phi_2 = -.65$, $\theta_1 = -0.9$, $\theta_2 = 0.79$, $F(y_{t-d}; \gamma, c)$ equal to the logistic function (2) [LSTAR] or the exponential function [ESTAR] with $d = 1$, $\gamma = 10$, and $\sigma = 0.2$. The table is based on 1000 replications of length $T = 300$.

Table 6: Behavior of decision rules in the presence of additive outliers[1]

θ_0	c	π	ξ	LSTAR TP	LSTAR RTP	LSTAR EJP	LSTAR REJP	ESTAR TP	ESTAR RTP	ESTAR EJP	ESTAR REJP
0.00	0.00	.00	0	0.997	1.000	1.000	0.966	0.977	0.954	0.998	0.831
		.05	3	0.498	0.917	0.882	0.871	0.928	0.879	0.995	0.750
			5	0.151	0.905	0.306	0.825	0.927	0.659	0.999	0.659
			7	0.080	0.984	0.088	0.863	0.941	0.791	0.999	0.607
0.02	0.00	.00	0	1.000	0.999	1.000	0.957	0.977	0.943	0.998	0.827
		.05	3	0.519	0.920	0.854	0.848	0.930	0.877	0.991	0.720
			5	0.146	0.909	0.230	0.799	0.927	0.651	0.997	0.631
			7	0.074	0.983	0.063	0.875	0.944	0.778	1.000	0.605
0.04	0.02	.00	0	0.995	0.999	1.000	0.966	0.972	0.942	0.995	0.800
		.05	3	0.485	0.915	0.874	0.855	0.927	0.862	0.992	0.716
			5	0.145	0.906	0.298	0.825	0.927	0.669	0.997	0.631
			7	0.079	0.980	0.084	0.835	0.941	0.778	0.998	0.595

[1] Relative frequencies of correctly choosing the correct type of model from the decision rules described in Section 2.2. Frequencies are conditional on rejecting the null hypothesis of linearity by the general test corresponding to the decision rules (e.g. the entries in the column EJP are conditional on rejection of linearity by test NL4). Series are generated according to (8) and (7) with $\phi_0 = 0$, $\phi_1 = 1.4$, $\phi_2 = -.65$, $\theta_1 = -0.9$, $\theta_2 = 0.79$, $F(y_{t-d}; \gamma, c)$ equal to the logistic function (2) or the exponential function (3) with $d = 1$, $\gamma = 10$, and $\sigma = 0.2$. The table is based on 1000 replications of length $T = 300$.

Table 7: Regular and outlier robust specification procedures for STAR models for monthly gold returns[1]

	Test	d = 1	2	3	4	5	6
				Regular			
	NL3	0.001	0.042	0.001	0.002	0.000	0.013
TP	H_{03}	0.169	0.211	0.200	0.020	0.039	0.121
	H_{02}	0.000	0.290	0.003	0.117	0.000	0.903
	H_{01}	0.980	0.023	0.037	0.492	0.103	0.002
	NL4	0.001	0.038	0.000	0.002	0.000	0.013
EJP	H_{0L}	0.002	0.314	0.002	0.009	0.033	0.018
	H_{0E}	0.582	0.068	0.000	0.003	0.000	0.388
				Robust			
	NL3	0.018	0.491	0.620	0.568	0.245	0.767
TP	H_{03}	0.444	0.800	0.347	0.004	0.000	0.199
	H_{02}	0.486	0.479	0.796	0.019	0.016	0.779
	H_{01}	0.004	0.118	0.611	0.851	0.749	0.730
	NL4	0.041	0.680	0.664	0.568	0.245	0.752
EJP	H_{0L}	0.375	0.680	0.155	0.081	0.020	0.405
	H_{0E}	0.275	0.130	0.008	0.109	0.400	0.206

[1] p-values for LM-type tests against smooth transition nonlinearity and tests of sub-hypothesis in the specification procedures of Teräsvirta (1994) [TP] and Escribano and Jordá (1997) [EJP] for monthly returns on gold. The upper panel gives p-values for standard tests, the lower panel for LM-type tests which are robust to additive outliers. The various null hypotheses are given in Sections 2.1 and 2.2.

Table 8: Regular and outlier robust specification procedures for STAR models for monthly silver returns[1]

	Test	d = 1	2	3	4	5	6
				Regular			
	NL3	0.005	0.000	0.000	0.000	0.001	0.004
TP	H_{03}	0.081	0.028	0.002	0.001	0.005	0.146
	H_{02}	0.006	0.000	0.000	0.017	0.763	0.689
	H_{01}	0.199	0.281	0.000	0.011	0.002	0.001
	NL4	0.005	0.000	0.000	0.000	0.001	0.000
	H_{0L}	0.068	0.005	0.001	0.000	0.280	0.289
	H_{0E}	0.004	0.039	0.036	0.000	0.205	0.001
				Robust			
	NL3	0.939	0.168	0.004	0.325	0.086	0.411
TP	H_{03}	0.400	0.496	0.868	0.144	0.000	0.555
	H_{02}	0.542	0.302	0.002	0.721	0.495	0.253
	H_{01}	0.947	0.429	0.000	0.664	0.863	0.116
	NL4	0.463	0.282	0.005	0.488	0.086	0.470
EJP	H_{0L}	0.636	0.222	0.490	0.153	0.013	0.300
	H_{0E}	0.386	0.035	0.652	0.066	0.686	0.043

[1] p-values for LM-type tests against smooth transition nonlinearity and tests of sub-hypothesis in the specification procedures of Teräsvirta (1994) [TP] and Escribano and Jordá (1997) [EJP] for monthly returns on silver. The upper panel gives p-values for standard tests, the lower panel for LM-type tests which are robust to additive outliers. The various null hypotheses are given in Sections 2.1 and 2.2.

Table 9: p-values of LM-type tests for real exchange rate data vis-a-vis US dollar, 1973.01-1994.12[1]

		Test			
Country	p	NL3	RNL3	NL4	RNL4
Australia	12	0.272	0.792	0.199	0.731
Belgium	3	0.111	0.159	0.089	0.137
Canada	12	0.037	0.286	0.050	0.249
Denmark	12	0.077	0.099	0.050	0.333
Finland	12	0.005	0.105	0.001	0.206
France	1	0.023	0.107	0.029	0.189
Germany	1	0.104	0.298	0.043	0.145
Greece	13	0.028	0.034	0.025	0.048
Italy	12	0.182	0.073	0.082	0.018
Japan	13	0.142	0.379	0.095	0.272
Luxemburg	12	0.224	0.226	0.264	0.199
The Netherlands	12	0.072	0.142	0.145	0.240
Norway	7	0.001	0.224	0.003	0.452
Portugal	13	0.449	0.390	0.301	0.259
Spain	8	0.109	0.733	0.048	0.356
Sweden	10	0.028	0.098	0.012	0.206
Switzerland	12	0.021	0.320	0.023	0.070
United Kingdom	12	0.053	0.141	0.036	0.133

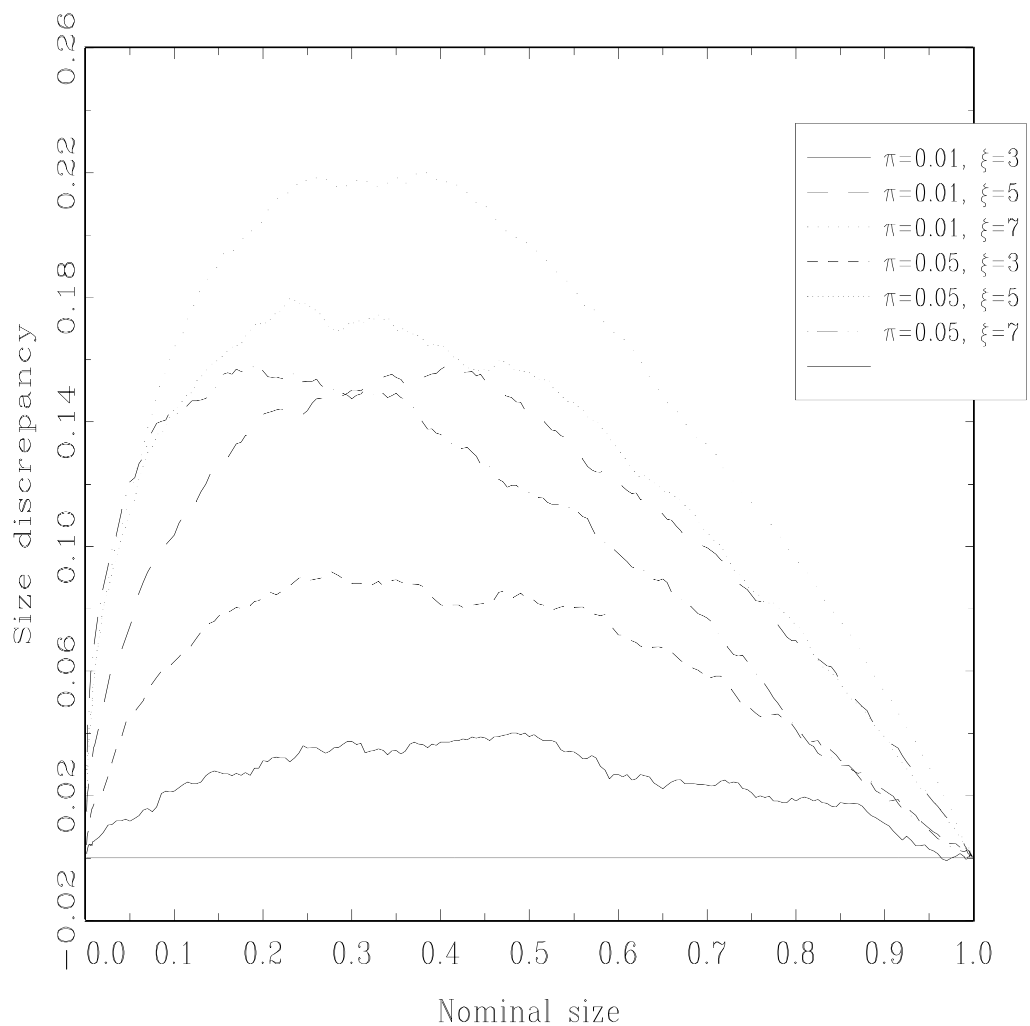
[1] Minimum p-values for LM-type tests against smooth transition nonlinearity in monthly real exchange rates vis-a-vis the US dollar, based on consumer price indices, January 1973-December 1994. Tests are computed for $d = 1, \ldots, 6$. The order p of the linear model under the null hypothesis is chosen according to the procedure of Ng and Perron (1995) outlined in Section 5.2.

Table 10: Model selection for real exchange rate data, 1973.01-1994.12[1]

	Decision rule			
Country	TP	RTP	EJP	REJP
Australia	Linear	Linear	Linear	Linear
Belgium	Linear	Linear	LSTAR(1)	Linear
Canada	LSTAR(4)	Linear	LSTAR(4)	Linear
Denmark	ESTAR(6)	LSTAR(6)	LSTAR(1)	Linear
Finland	LSTAR(1)	Linear	LSTAR(1)	Linear
France	ESTAR(5)	Linear	ESTAR(1)	Linear
Germany	Linear	Linear	LSTAR(1)	Linear
Greece	LSTAR(6)	ESTAR(1)	LSTAR(6)	LSTAR(1)
Italy	Linear	LSTAR(6)	LSTAR(6)	LSTAR(6)
Japan	Linear	Linear	ESTAR(1)	Linear
Luxemburg	Linear	Linear	Linear	Linear
The Netherlands	LSTAR(1)	Linear	Linear	Linear
Norway	ESTAR(1)	Linear	LSTAR(1)	Linear
Portugal	Linear	Linear	Linear	Linear
Spain	Linear	Linear	LSTAR(3)	Linear
Sweden	LSTAR(2)	ESTAR(5)	LSTAR(2)	Linear
Switzerland	ESTAR(2)	Linear	LSTAR(2)	LSTAR(6)
United Kingdom	LSTAR(4)	Linear	ESTAR(4)	Linear

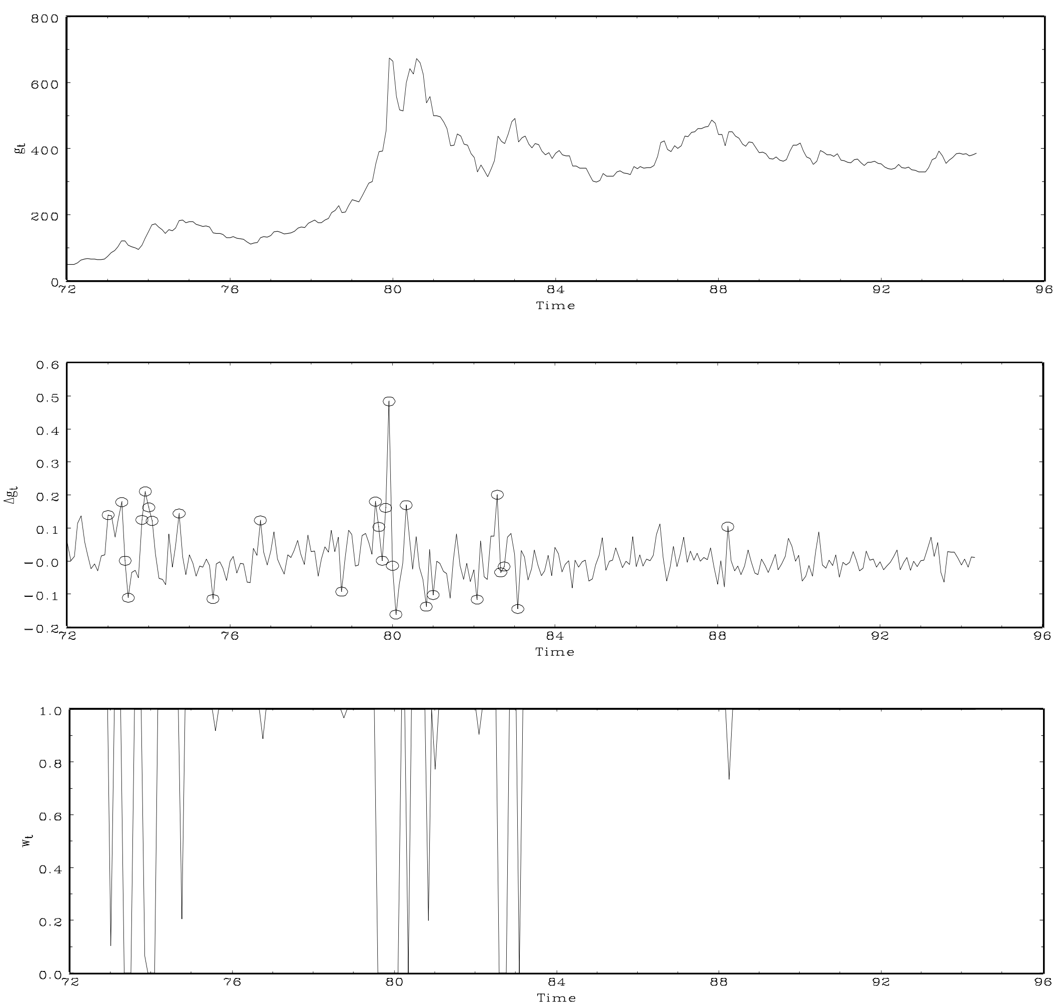
[1] Model selection for monthly real exchange rates vis-a-vis the US dollar, based on wholesale price indices, January 1973-December 1994 by decision rules of Teräsvirta (1994) [TP] and Escribano and Jordá (1997) [EJP], as well as their robust counterparts (RTP and REJP, respectively). The choice for the delay parameter d is given in parentheses. Decision rules for which the p-value of the corresponding general linearity test given in Table 9 is larger than 0.10 are denoted as 'Linear'.

Figure 1: p value discrepancy plots for NL3 test, DGP I



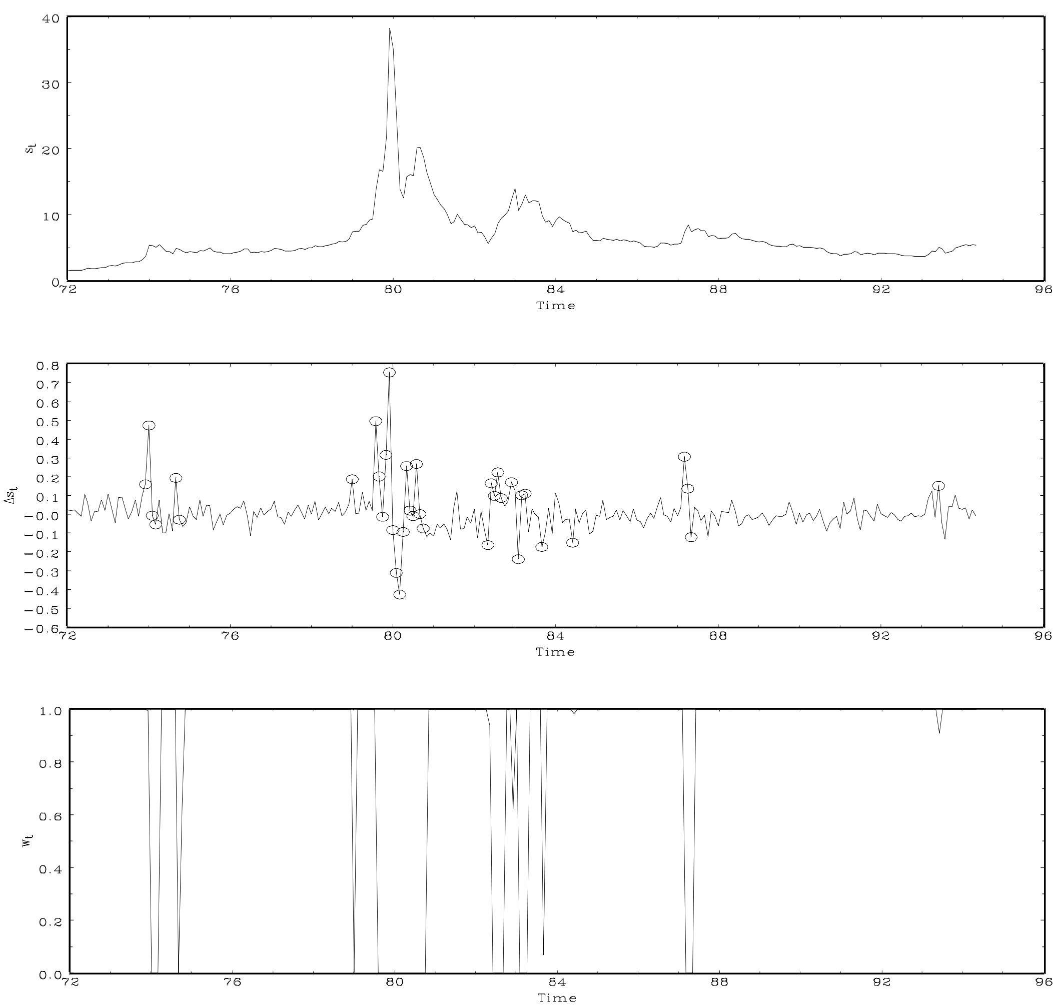
Note: p value discrepancy plots for NL3 test, DGP I. The graph shows the difference between the actual rejection frequency and the asymptotic nominal significance level $(\hat{F}(x_i) - x_i)$ versus the nominal significance level x_i. The $\hat{F}(x_i)$ have been calculated for $x_i = .001, .002, \ldots, .010, .015, \ldots, .990, .991, \ldots, .999$ ($M = 215$). Series are generated according to (6) and (7) with $\phi = (0, .2, .2)'$ and $\sigma = 1$. The graph is based on 5000 replications for sample size $T = 300$.

Figure 2: Gold prices



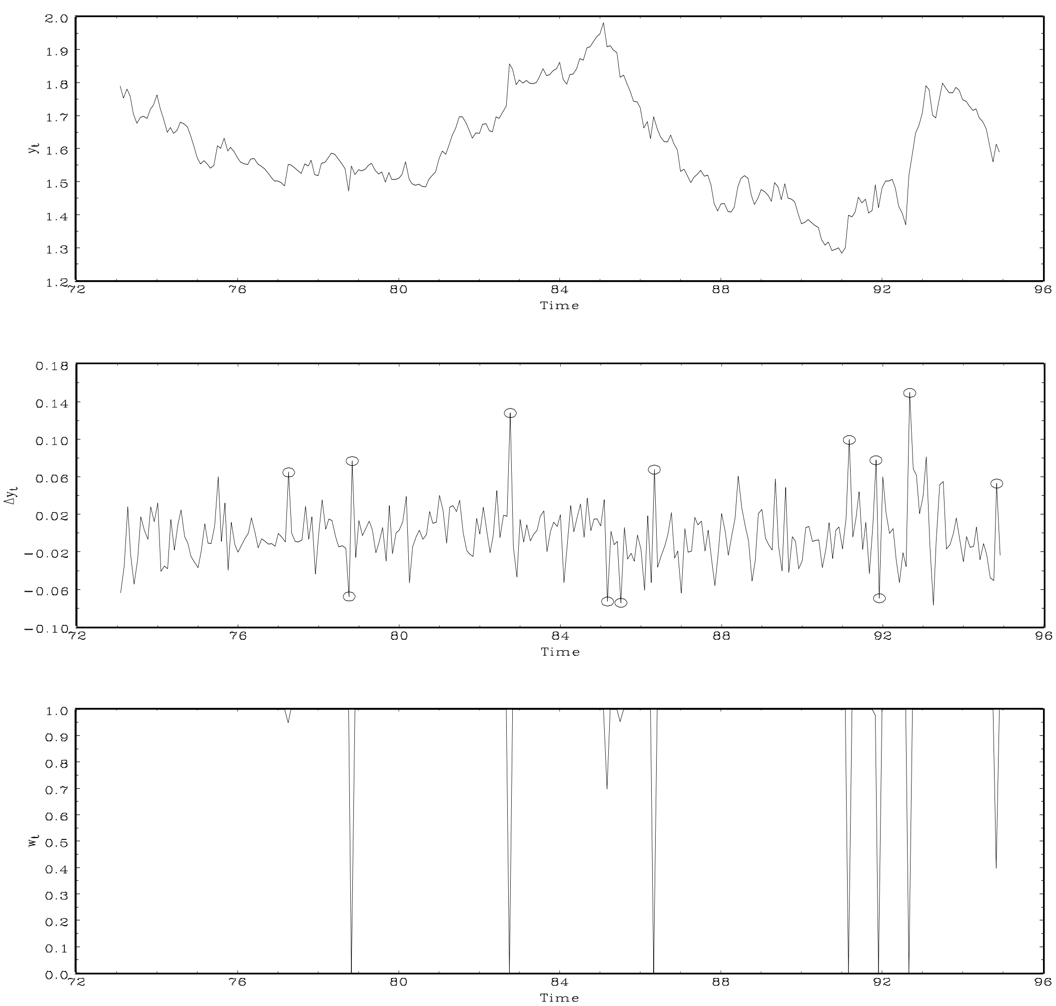
Note: The upper panel shows monthly gold prices in US$ per fine ounce in London, January 1971-June 1994. The middle panel shows the corresponding returns. Observations which receive a weight less than 1 in the GM estimation procedure for the parameters in the AR(2) model are marked with open circles. The actual weights are graphed in the lower panel.

Figure 3: Silver prices



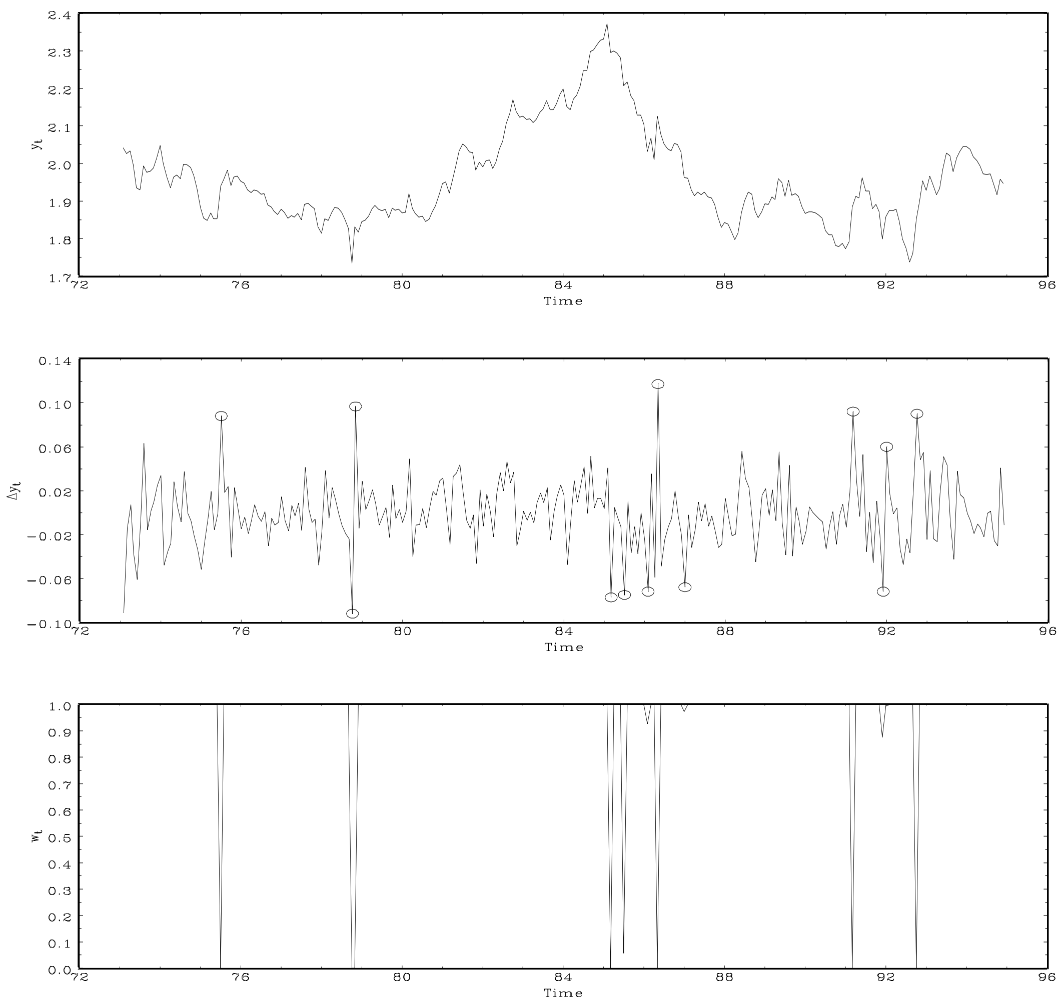
Note: The upper panel shows monthly silver prices in US$ per troy ounce in New York, January 1971-June 1994. The middle panel shows the corresponding returns. Observations which receive a weight less than 1 in the GM estimation procedure for the parameters in the AR(2) model are marked with open circles. The actual weights are graphed in the lower panel.

Figure 4: Real exchange rate - Finnish markka



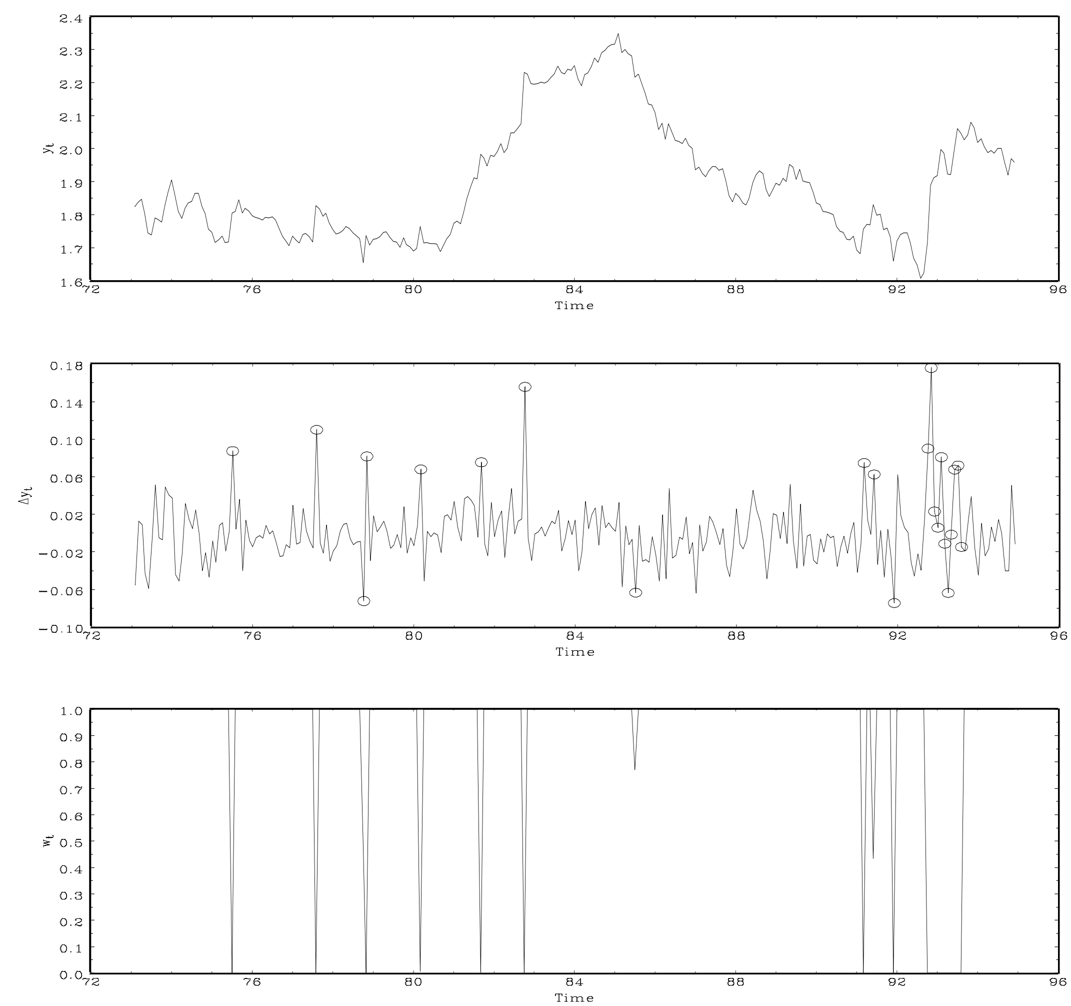
Note: The upper panel shows the log monthly real exchange rate of the Finnish markka vis-à-vis the US$, January 1973-December 1994. The middle panel shows the corresponding returns. Observations which receive a weight less than 1 in the GM estimation procedure for the parameters in the AR(p) model are marked with open circles. The actual weights are graphed in the lower panel.

Figure 5: Real exchange rate - Norwegian kroner



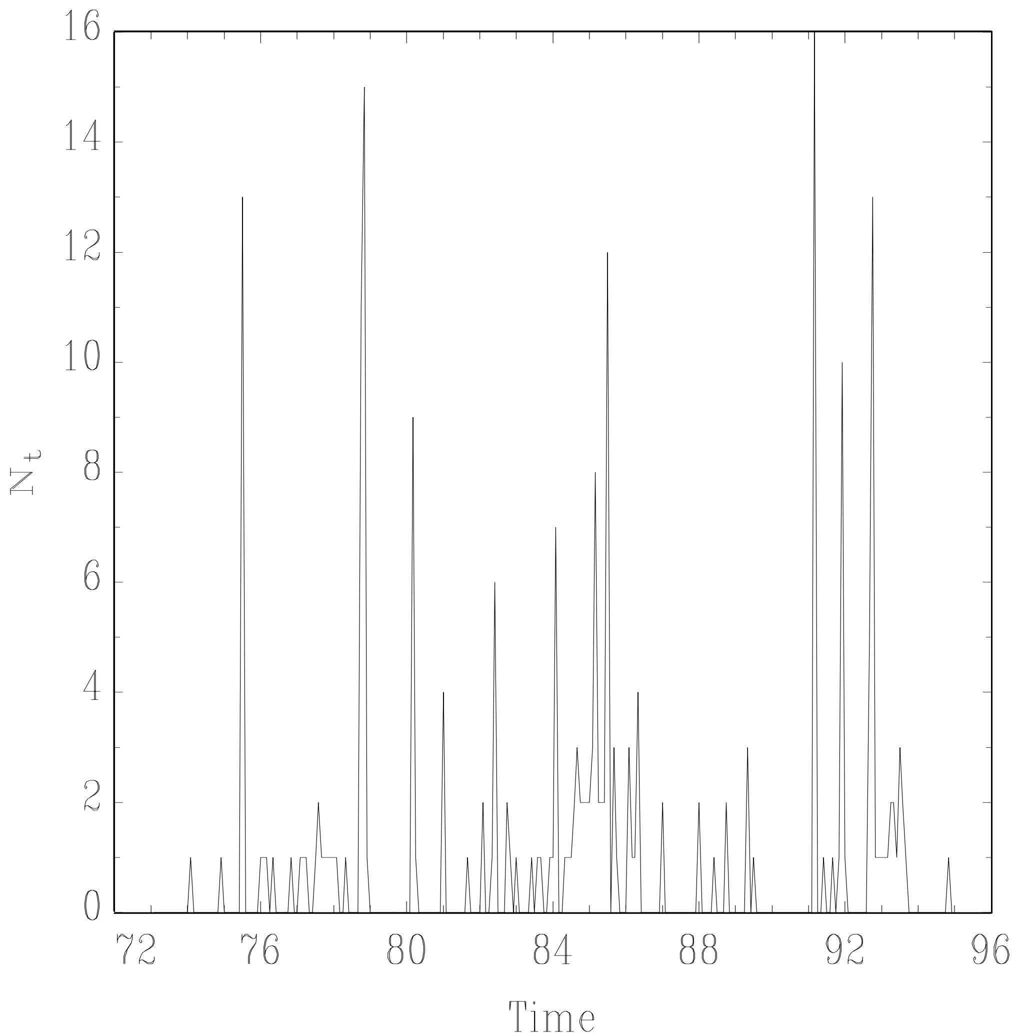
Note: The upper panel shows the log monthly real exchange rate of the Norwegian kroner vis-à-vis the US$, January 1973-December 1994. The middle panel shows the corresponding returns. Observations which receive a weight less than 1 in the GM estimation procedure for the parameters in the AR(p) model are marked with open circles. The actual weights are graphed in the lower panel.

Figure 6: Real exchange rate - Swedish kroner



Note: The upper panel shows the log monthly real exchange rate of the Swedish kroner vis-à-vis the US$, January 1973-December 1994. The middle panel shows the corresponding returns. Observations which receive a weight less than 1 in the GM estimation procedure for the parameters in the AR(p) model are marked with open circles. The actual weights are graphed in the lower panel.

Figure 7: Outliers in real exchange rates



Note: The figure shows the number of countries for which the respective observations on the real exchange rate vis-à-vis the US$ receive a weight less than 1 in the GM estimation procedure.