Testing for Smooth Transition Nonlinearity in the Presence of Outliers

Dick van Dijk*
Tinbergen Institute, Erasmus University Rotterdam

Philip Hans Franses†
Rotterdam Institute for Business Economic Research and
Econometric Institute, Erasmus University Rotterdam

André Lucas
Financial Sector management, Free University Amsterdam

First version: February 21, 1996
This version: May 17, 1996

Abstract

Regime-switching models, like the smooth transition autoregressive (STAR) model are typically applied to time series of moderate length. Hence, the nonlinear features which these models intend to describe may be reflected in only a few observations. Conversely, neglected outliers in a linear time series of moderate length may incorrectly suggest STAR type nonlinearity. In this paper we propose outlier robust tests for STAR type nonlinearity. These tests are designed such that they have a better level and power behavior than standard nonrobust tests in situations with outliers. We formally derive local and global robustness properties of the new tests. Extensive Monte Carlo simulations show the practical usefulness of the robust tests. An application to several quarterly industrial production indices illustrates that apparent nonlinearity in time series sometimes seems due to only a small number of outliers.

Keywords: Nonlinearity, Outliers, Robust estimation

*Correspondence to Dick van Dijk, Tinbergen Institute, Burg. Oudlaan 50, 3062 PA Rotterdam, The Netherlands, Email: dvandijk@tir.few.eur.nl

†Financial support from the Royal Netherlands Academy of Arts and Sciences is gratefully acknowledged.

1 Introduction

Much recent research in both theoretical and applied time series analysis has focused on nonlinear features of economic variables. Typical nonlinear time series models which appear useful in practice concern various forms of regime-switches. Examples are the Markov regime-switching model of Hamilton (1989), the threshold autoregressive (TAR) model, discussed in Tong (1990) and Tsay (1989), and the smooth transition autoregressive (STAR) models, advocated by Teräsvirta (1994), Granger and Teräsvirta (1993) and others. These models have been applied most frequently to study possible nonlinearity of business cycles, see Luukkonen and Teräsvirta (1991) and Teräsvirta and Anderson (1992). Most macroeconomic variables which are thought to measure the business cycle are sampled only quarterly or annually. Therefore, only series of moderate length are available and, consequently, it may be that possible nonlinear properties are reflected in only a small number of observations. One may then be tempted to view these 'nonlinear data points' as aberrant observations and remove them using some of the familiar outlier removal techniques, see Balke and Fomby (1994), *inter alia*. This might even be justified by noting that nonlinear time series models typically involve many additional parameters and one may want to prevent estimating these parameters for only a few observations. However, removing outliers too drastically may accidentally destroy intrinsic nonlinearity, which, for example, could have been exploited for forecasting. Conversely, in case of a linear time series which is contaminated by outliers, nonlinearity tests may point towards nonlinear structures, which in turn can lead to estimating 'too' complicated models. In sum, there seems a need for modelling strategies and tests which are capable of distinguishing between nonlinearity and outliers.

In the present paper we confine ourselves to proposing tests for smooth transition nonlinearity in the presence of outliers. We consider outlier robust estimation techniques, see, e.g., Huber (1981), Martin (1981) and Hampel *et al.* (1986), to modify the tests developed by Luukkonen *et al.* (1988). The outline of this paper is as follows. In section 2, we briefly review some outlier models which are considered relevant in the time series literature. We also present the effect of outliers on ordinary least squares estimates of linear time series models. Robust estimation methods for time series models are considered in section 3. In section 4 we discuss tests for STAR nonlinearity. At the end of this section, we propose robustified versions of the test statistics. The effect of outliers on these nonlinearity tests is investigated analytically in section 5, while the empirical size and power properties are evaluated by means of Monte Carlo experiments in section 6. We find that our robust testing procedure works remarkably well. An empirical illustration is provided in section 7, where the tests are applied to various industrial production series, previously analysed in Luukkonen and Teräsvirta (1991), Teräsvirta and Anderson (1992) and Teräsvirta *et al.* (1994). The general outcome is that we find similar nonlinear features as in those studies, except for three series for which apparent nonlinearity appears to be due to a small number of observations. Finally, in section 8 we conclude with some remarks and suggestions for further research.

2 A brief discussion on outliers

Davies and Gather (1993) note that outliers are always defined with respect to a model. Certain observations might be outliers in one model and at the same time be perfectly regular observations in another model. Most of the literature on outlier detection and estimation in the presence of outliers has concentrated on linear time series models. In this section, we briefly consider some aspects which will prove to be relevant for nonlinearity testing later on.

A useful starting point for a brief discussion on outliers in time series is the replacement model of Martin and Yohai (1986),

$$y_t = x_t(1-\delta_t) + \zeta_t\delta_t\ , \qquad t = 1,\ldots,T\ , \tag{1}$$

where δ_t is a binary random variable which equals 1 with probability π and 0 otherwise and where T denotes the sample size. The observed time series y_t consists of a *core* process x_t and a *contaminating* process ζ_t. In the following, we assume that x_t is governed by an autoregressive (AR) process of order p, i.e., $\phi(B)x_t = \varepsilon_t$, where $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ is a polynomial in the backshift operator B, $B^k x_t = x_{t-k}$, and where $\varepsilon_t \sim$ i.i.d.$(0, \sigma_\varepsilon^2)$. Different specifications of the ζ_t process can generate a wide variety of outlier patterns.

Two types of contamination are usually considered to be of special interest in the analysis of outlier effects on time series. First, an additive outlier (AO) model is obtained if $\zeta_t = x_t + \zeta$ for some constant ζ, such that (1) reduces to

$$y_t = x_t + \zeta\delta_t \;, \qquad t = 1, \ldots, T \;. \tag{2}$$

An innovative outlier (IO) model results if $\zeta_t = x_t + \zeta/\phi(B)$, which yields

$$y_t = x_t + \zeta/\phi(B)\delta_t \;, \qquad t = 1, \ldots, T \;. \tag{3}$$

The AO case gives a one time effect on the level of the time series, as only the current observation y_t is affected. In the IO model, however, a shock at time t also influences future observations $y_{t+1}, y_{t+2}, \ldots$, through the same dynamics as the linear part of the model.

Hoek *et al.* (1995) investigate the effect of both isolated AO's and IO's at time $t = s$ when the core process x_t follows an AR(1) model

$$x_t = \phi x_{t-1} + \varepsilon_t \;, \qquad t = 1, \ldots, T \;, \tag{4}$$

with $|\phi| < 1$. It is shown that an AO corresponds with two outliers in the (y_{t-1}, y_t) plane. Using the classification of Rousseeuw and van Zomeren (1990), the point (y_{s-1}, y_s) is a vertical outlier as y_s falls outside the range of the majority of the data. The next point (y_s, y_{s+1}) is a so-called bad leverage point, characterized by an abnormal value of the regressor. In case of an IO on the other hand, the vertical outlier at $t = s$ is followed by a number of good leverage points, which are characterized by large values for both y_{t-1} and y_t that approximately satisfy the linear AR(1) model.

Denby and Martin (1979), Bustos and Yohai (1986) and Martin and Yohai (1986), among others, consider the estimation of the parameters of AR models in the presence of outliers. In the presence of IO's, the ordinary least squares (OLS) estimates of the autoregressive parameters are consistent, although they are inefficient. AO's have a much more disastrous effect on the OLS estimates. If, for example, the core process x_t follows the AR(1) model (4) and the observed time series y_t is contaminated with an isolated additive outlier of magnitude ζ, the OLS estimator of ϕ calculated with the observed series approaches zero as ζ tends to inifinity. In general, if the probability of occurence of an AO of size ζ is equal to π, the OLS estimator of ϕ is biased towards zero with probability limit given by

$$\operatorname*{plim}_{T\to\infty} \hat{\phi} = \frac{\phi}{1 + \frac{\pi\,\zeta^2}{\sigma_x^2}}, \tag{5}$$

where σ_x^2 denotes the variance of x_t, $\sigma_x^2 = \sigma_\varepsilon^2/(1-\phi^2)$.

Over the years, a number of outlier detection and correction procedures have been developed, see, e.g., Tsay (1986*a*), Chang *et al.* (1988) and Chen and Liu (1993). All these procedures are characterized by an iterative 'estimation-detection-correction-estimation' scheme, which may make them subjective and time-consuming. An alternative method to cope with outliers is to use robust estimation techniques. These latter techniques are the subject of the next section.

3 Robust estimation methods for time series models

In this section we highlight some issues in robust estimation which are needed to modify tests for STAR nonlinearity in the presence of outliers. For a more general discussion we refer to Denby and Martin (1979) and Martin (1981). Consider again the simple AR(1) model for the uncontaminated series x_t,

$$x_t = \phi x_{t-1} + \varepsilon_t \ , \quad t = 1, \ldots, T \ , \tag{6}$$

where $|\phi| < 1$ and where the errors ε_t are again assumed to be independent white noise. Furthermore, assume that the observed time series y_t follows the replacement model (1) for a general contamination process ζ_t. The OLS estimate of the autoregressive parameter ϕ based on the observed series minimizes the sum of squared residuals, which can equivalently be characterized by the first order condition

$$\sum_{t=1}^{T} y_{t-1}(y_t - \phi y_{t-1}) = 0 \ . \tag{7}$$

In order to avoid the deficiencies of the OLS estimator discussed in the previous section, the autoregressive parameter can be estimated robustly using maximum likelihood type (M) or Generalized M (GM) estimators. The class of GM estimators is designed to estimate ϕ giving less weight to influential observations such as leverage points and vertical outliers. Here we consider the Schweppe type of GM estimators (Handschin *et al.* (1975)), which solves the alternative first order condition

$$\sum_{t=1}^{T} w_y(y_{t-1}) y_{t-1} \cdot \psi\left(\frac{y_t - \phi y_{t-1}}{\sigma_\varepsilon w_y(y_{t-1})}\right) = 0 \ , \tag{8}$$

where $\psi(\cdot)$ is an odd and bounded function and $w_y(\cdot)$ is a weight function that assigns weights between 0 and 1 to the regressor y_{t-1}. The function $\psi(\cdot)$ must satisfy certain additional regularity conditions in order for the GM estimator to be consistent and asymptotically normal, see Hampel *et al.* (1986).

If the regressors are not weighted, i.e., $w_y(y_{t-1}) = 1$ for all t, the GM estimators reduce to M estimators, while the usual OLS estimator is obtained if, in addition, $\psi(r) = r$. Denby and Martin (1979) show that in the presence of IO's M estimators are efficient, while the asymptotic variance of GM estimators is larger due to the weighting of the regressors. Both M and GM estimators are asymptotically biased in an AO setting, although, if $w_y(\cdot)$ and $\psi(\cdot)$ are chosen properly, the bias of the

GM estimator can be considerably smaller. As AO's have the most damaging effect on the OLS estimator, we will focus on the GM estimator only. We further focus on the Schweppe form of the GM estimator, because this estimator only downweights vertical outliers and bad leverage points, while it fully exploits the correct signal in good leverage points, see Hampel *et al.* (1986).

Defining $w_r(r) = \psi(r)/r$ for $r \neq 0$ and $w_r(0) = 1$, the first order condition (8) can be rewritten as

$$\sum_{t=1}^{T} y_{t-1} \cdot w_r[(y_t - \phi y_{t-1})/(\sigma_\varepsilon w_y(y_{t-1}))](y_t - \phi y_{t-1}) = 0 \; , \tag{9}$$

from which it can be inferred that the GM estimator is a type of weighted least squares estimator. The weight for the t-th observation is given by the value of $w_r(\cdot)$, which depends upon the unknown parameter ϕ. The functions $w_y(\cdot)$ and $\psi(\cdot)$ now should be chosen such that the t-th observation receives a relatively small weight if either y_{t-1} or $(y_t - \phi y_{t-1})/\sigma_\varepsilon$ becomes large (in absolute value).

Common choices for the $\psi(\cdot)$ function in (8) are the Huber function and the Tukey bisquare function. The Huber ψ function is given by $\psi(r) = \mathrm{med}(-c, c, r)$, where med denotes the median and $c > 0$. The tuning constant c determines the robustness of the estimator to outliers and the efficiency of the estimator in the absence of outliers. The robustness of the estimator is decreasing in c, while the efficiency in the absence of outliers increases with c. Therefore, the tuning constant c should be chosen such that these two properties of the estimator are balanced. Usually, c is taken equal to 1.345 to produce an estimator that has an efficiency of 95% if ε_t is normally distributed. The weights $w_r(\cdot)$ that result from applying the Huber function have the attractive property that they equal 1 if $r \in (-c, c)$. Only observations for which the (standardized) residual is outside this region receive less weight. A disadvantage is that these weights decline to zero only very slowly. Thus, subjective judgement is required to decide whether a weight is small or not.

The Tukey bisquare function is given by $\psi(r) = r(1 - (r/c)^2)^2 \cdot (1 - H(|r| - c)$, where $H(\cdot)$ is the Heaviside function, $H(z) = 1$ if $z > 0$ and $H(z) = 0$ if $z \leq 0$. The same considerations apply with respect to the tuning constant c as for the Huber function. Usually c is set equal to 4.685, again to achieve 95% efficiency for normally distributed ε_t. The Tukey function might be regarded as the mirror-image of the Huber ψ in that downweighting occurs for all nonzero values of r and the resulting weights decline to zero quite rapidly.

In this paper we use the polynomial ψ function as proposed in Lucas *et al.* (1996), given by

$$\psi(r) = r(1 - H(|r| - c_1))\text{sgn}(r) + H(|r| - c_1)(1 - H(|r| - c_2))g(|r|) \; , \tag{10}$$

where c_1 and c_2 are tuning constants, sgn is the signum function, and $g(|r|)$ is a fifth order polynomial such that $\psi(r)$ is twice continuously differentiable. This ψ function combines the attractive properties of the Huber and Tukey functions. Observations receive a weight equal to 1 if their standardized residuals are within $(-c_1, c_1)$ and a weight equal to zero if the residuals are larger than c_2 in absolute value. Partial weighting occurs in-between. The tuning constants c_1 and c_2 are taken to be the square roots of the 0.99 and 0.999 quantiles of the $\chi^2(1)$ distribution, that is, $c_1 = 2.576$ and $c_2 = 3.291$. The weights $w_r(\cdot)$ resulting from the three ψ functions discussed above are graphed in Figure 1, which clearly reveals the differences and similarities between the different functions.

The weight function $w_y(\cdot)$ for the regressor is specified as

$$w_y(y_{t-1}) = \psi(d(y_{t-1})^\alpha)/d(y_{t-1})^\alpha \; , \tag{11}$$

where again $\psi(\cdot)$ is given by (10), $d(y_{t-1})$ is the Mahalanobis distance of y_{t-1}, i.e., $d(y_{t-1}) = |y_{t-1} - m_y|/\sigma_y$, with m_y and σ_y measures of location and scale of y_{t-1}, respectively. These measures are estimated robustly by the median and the median absolute deviation (MAD), respectively, i.e., $m_y = \text{med}(y_{t-1})$ and $\sigma_y = 1.483 \cdot \text{med}|y_{t-1} - m_y|$. The constant 1.483 is used to make the MAD a consistent estimator of the standard deviation in case ε_t is normally distributed. Finally, following Simpson *et al.* (1992), the constant α in (11) is set equal to 2 in order to obtain robustness of standard errors.

The first order condition (9) is nonlinear in ϕ and, therefore, estimation requires an iterative procedure. In fact, interpreting $w_r(\cdot)$ as a function of $(\phi, \sigma_\varepsilon)$, $w_r(\phi, \sigma_\varepsilon)$, and denoting the estimates of ϕ and σ_ε at the n-th iteration by $\hat{\phi}^{(n)}$ and $\hat{\sigma}_\varepsilon^{(n)}$, respectively, it follows from (9) that $\hat{\phi}^{(n+1)}$ might be computed as the weighted least squares estimate

$$\hat{\phi}^{(n+1)} = \left(\sum_{t=1}^{T} w_r(\hat{\phi}^{(n)}, \hat{\sigma}_\varepsilon^{(n)}) y_{t-1} y_t \right) / \left(\sum_{t=1}^{T} w_r(\hat{\phi}^{(n)}, \hat{\sigma}_\varepsilon^{(n)}) y_{t-1}^2 \right) , \qquad (12)$$

where the estimate of σ_ε can be updated at each iteration using the MAD estimator given above.

In order to have maximum protection against outliers, the breakdown point of the estimator, that is, the maximum fraction of contaminated observations the estimator can cope with, should be as high as possible. We follow Simpson *et al.* (1992) and Coakley and Hettmansperger (1993), who show that if a high breakdown point (HBP) estimator is used to construct starting values and if only one iteration according to (12) is performed, an efficient estimator is obtained which retains the high breakpoint of the initial estimator. We use the least median of squares (LMS) estimator of Rousseeuw (1984) to obtain a starting value for the autoregressive parameter, $\hat{\phi}^{(0)}$, and we apply the MAD estimator to the corresponding residuals to obtain an initial scale estimate, $\hat{\sigma}^{(0)}$. The resulting HBP-GM estimator for ϕ has a breakdown point of approximately 0.5. In the next section we will use this robust estimator to modify tests for STAR nonlinearity.

4 Smooth transition nonlinearity

Consider the general STAR model of order p [STAR(p)] for a univariate time series y_t,

$$y_t = \phi' y_t^{(p)} + f(\tilde{y}_t^{(p)}; \gamma, a, c)\theta' y_t^{(p)} + \varepsilon_t \ , \quad t = 1, \ldots, T \ , \tag{13}$$

where $y_t^{(p)} = (1, \tilde{y}_t^{(p)})'$, $\tilde{y}_t^{(p)} = (y_{t-1}, \ldots, y_{t-p})'$, $\phi = (\phi_0, \phi_1, \ldots, \phi_p)'$ and $\theta = (\theta_0, \theta_1, \ldots, \theta_p)'$. For a logistic STAR (LSTAR) model, the transition function $f(\tilde{y}_t^{(p)}; \gamma, a, c)$ is taken to be the logistic function

$$f(\tilde{y}_t^{(p)}; \gamma, a, c) = (1 + \exp\{-\gamma(a'\tilde{y}_t^{(p)} - c)\})^{-1}, \quad \gamma > 0 \ , \tag{14}$$

where $a = (a_1, \ldots, a_p)'$, while γ and c are scalars. The LSTAR model is member of the class of regime-switching models in that the time series can move between two extreme regimes, associated with the values 0 and 1 for the transition function $f(\tilde{y}_t^{(p)}; \gamma, a, c)$, where the transition between these two regimes is smooth and governed by lagged values of the time series itself. For an extensive discussion of STAR models we refer to Teräsvirta (1994) and Granger and Teräsvirta (1993).

4.1 Nonrobust tests for smooth transition nonlinearity

In this paper we consider the Lagrange Multiplier (LM) tests to identify LSTAR-type nonlinearity developed by Luukkonen *et al.* (1988). The null hypothesis of linearity can be taken as $H_0 : \theta = 0$ in (13). This hypothesis is tested against the alternative $H_1 : \theta \neq 0$. It is immediately seen from (13) with (14) that the model is not identified under the null hypothesis, because if H_0 holds, γ, a and c can take any value. Consequently, the usual asymptotic theory cannot be applied to derive LM tests, see Davies (1977,1987) for a general discussion of this identification problem. Luukkonen *et al.* (1988) suggest to remedy this problem by replacing the transition function $f(\tilde{y}_t^{(p)}; \gamma, a, c)$ in (13) by a suitable approximation. In the resulting auxiliary model, the identification problem is no longer present and linearity can easily be tested by so-called LM-type tests.

In general, the reparameterized model that is used for linearity testing can be written as

$$y_t = \phi' y_t^{(p)} + \theta^{*\prime} q_t + \varepsilon_t \,, \tag{15}$$

where ϕ and $y_t^{(p)}$ are defined above, q_t is an $m \times 1$ vector of auxiliary regressors containing higher-order and cross-product terms of the regressors in $\tilde{y}_t^{(p)}$, and θ^* is a parameter vector of corresponding length. The exact contents of the vector q_t depend on the approximation of the transition function chosen. Luukkonen *et al.* (1988) first propose to replace $f(\tilde{y}_t^{(p)}; \gamma, a, c)$ by a first-order Taylor approximation around the point $\gamma(a'\tilde{y}_t^{(p)} - c) = 0$. The corresponding statistic, denoted LM_1, is the general linearity test of Tsay (1986*b*), where the vector q_t consists of terms y_{t-i}^2 and $y_{t-i}y_{t-j}, i, j = 1, \ldots, p, i < j$. As this test might lack power against alternatives where only the intercept is changing (i.e. only $\theta_0 \neq 0$ in (13)), two alternative test statistics are put forward. The LM_2 statistic uses a third-order Taylor approximation for the transition function. The vector of auxiliary regressors in this case contains additional terms y_{t-i}^3, y_{t-i}^4, $y_{t-i}y_{t-j}^2$ and $y_{t-i}y_{t-j}^3$ for $i, j = 1, \ldots, p, i \neq j$. The last test, LM_3, is obtained from the LM_1 test by augmenting the relevant vector q_t with only the cubic terms y_{t-i}^3, as only these terms from the third-order approximation depend on the intercept θ_0.

For all three tests, the original null hypothesis of linearity, $H_0 : \theta = 0$, can be shown to be equivalent to the hypothesis that all coefficients of the auxiliary regressors gathered in q_t in (15) are zero, i.e. $H_0^* : \theta^* = 0$. The general form of the LM-type tests is then given by

$$\begin{aligned} LM_i &= \frac{\hat{\varepsilon}' Z (Z'Z)^{-1} Z' \hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}/T} \\ &= \frac{\hat{\varepsilon}'\hat{\varepsilon} - \hat{\varepsilon}'(I_T - Z(Z'Z)^{-1}Z')\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}/T} \,, \end{aligned} \tag{16}$$

$i = 1, 2, 3$, where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)'$ contains residuals estimated under the null hypothesis of linearity, $z_t = (y_t^{(p)\prime}, q_t')'$, $Z = (z_1', \ldots, z_T')'$, and I_T is the identity matrix of order T. Under a number of conditions, the LM_i statistic has an asymptotic $\chi^2(m)$

distribution, see Tsay (1986*b*) and Saikkonen and Luukkonen (1988). In small samples it is usually recommended to use an F version of the test. This version of the test can be computed as follows:

1. Estimate the model under the null hypothesis of linearity by regressing y_t on $y_t^{(p)}$. Compute the residuals $\hat{\varepsilon}_t$ and the sum of squared residuals $SSR_0 = \sum \hat{\varepsilon}_t^2$.

2. Perform the auxiliary regression of $\hat{\varepsilon}_t$ on $y_t^{(p)}$ and q_t and compute the sum of squared residuals from this regression, SSR_1.

3. The LM-type test statistic can now be computed as

$$LM_i = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T-m-p-1)}\ , \tag{17}$$

$i = 1, 2, 3$, which is approximately F distributed under the null hypothesis with m and $T-m-p-1$ degrees of freedom, respectively.

4.2 Robust tests for smooth transition nonlinearity

Because of the properties of the OLS estimator in the presence of AO's as discussed in section 2, one can expect that the LM_i tests can be severely affected by additive outliers. In the next section, we formally show that this is indeed the case. The robust estimators discussed in section 3 can be used to construct robust versions of the LM_i test statistics. In particular, we obtain a robust test by using a robust estimator to estimate the model under the null hypothesis. Hampel *et al.* (1986) and Peracchi (1991) show that the robustness properties of estimators carry over to test statistics based on these estimators. Moreover, under conventional assumptions, the LM_i tests retain their standard limiting χ^2 distributions. Thus, it might be expected that if the Schweppe form of the GM estimator is used for constructing LM-type test statistics, the resulting statistics are protected against the influence of (additive) outliers.

If we want to allow higher order models, i.e., $p > 1$, the robust HBP-GM estimator derived for the AR(1) case in section 3 needs to be generalized. This is fairly straightforward, except that now HBP estimators for multivariate location and scatter are required to compute the Mahalanobis distances in (11) for the regressors $y_t^{(p)}$. For this purpose, we use the minimum volume ellipsoid (MVE) estimator proposed by Rousseeuw (1985). The projection algorithm of Rousseeuw and van Zomeren (1990) is used to approximate this estimator.

By interpreting the first order condition for the AR-model given in (8) as a pseudo-score, an LM test can easily be constructed. Let $z_t = (y_t^{(p)\prime}, q_t')'$, $Z = (z_1, \ldots, z_T)'$, $\hat{W}_y = \text{diag}(\hat{w}_y(y_1^{(p)}), \ldots, \hat{w}_y(y_T^{(p)}))'$ and $\hat{\Psi} = (\hat{\psi}_r(r_1), \ldots, \hat{\psi}_r(r_T))'$, where $\hat{W}_y$ and $\hat{\Psi}$ are computed under the null hypothesis, and where r denotes the t-th standardized residual, $r_t \equiv (y_t - \phi y_{t-1})/(\sigma_\varepsilon w_y(y_{t-1}))$. The robust version of the LM-test statistics to test $H_0 : \theta^* = 0$ in (15), to be denoted as $RLM_i, i = 1, 2, 3$, can be computed as

$$\begin{aligned} RLM_i &= \frac{\hat{\Psi}'\hat{W}_y Z(Z'\hat{W}_y\hat{W}_y Z)^{-1}Z'\hat{W}_y\hat{\Psi}}{\hat{\Psi}'\hat{\Psi}/T} \\ &= \frac{\hat{\Psi}'\hat{\Psi} - \hat{\Psi}'(I - \hat{W}_y Z(Z'\hat{W}_y\hat{W}_y Z)^{-1}Z'\hat{W}_y)\hat{\Psi}}{\hat{\Psi}'\hat{\Psi}/T}, \end{aligned} \tag{18}$$

$i = 1, 2, 3$. Because $\psi(r) = w_r(r) \cdot r$, the term $\hat{\Psi}'\hat{\Psi} = \sum_{t=1}^T (\hat{w}_r(\hat{r}_t)\hat{\varepsilon}_t/\hat{\sigma}_\varepsilon \hat{w}_y(y_{t-1}^{(p)}))^2$ might be interpreted as a weighted sum of squared residuals under the null hypothesis, where the weights decline for large standardized residuals. As (18) shows, the F version of the test, corresponding to (17), can be computed by running an auxiliary OLS regression of the weighted residuals $\hat{\psi}(\hat{r}_t)$ on the weighted regressors $\hat{w}_y(y_t^{(p)})z_t$.

5 Outliers and tests for smooth transition nonlinearity

In this section we formally investigate the effect of outliers on the LM-type test statistics. We restrict our attention to the simple LM_1 and RLM_1 statistics when used to test an AR(1) model against an LSTAR(1) alternative. The qualitative results derived below remain the same for higher order models and for the other LM-type tests. We show that the presence of additive outliers leads to higher rejection rates for both the robust and nonrobust tests. The distortion for the nonrobust test, however, is much larger. If the outliers become extremely large or if the fraction of contamination becomes relatively high, the level of the nonrobust test is recovered, but the power of the test drops to its size. The power of the robust test, in contrast, is significantly higher.

In the following theorem we derive both a global and a local (non)robustness result for the RLM_1 test, based on an M estimator, i.e., $w_y(\cdot) \equiv 1$ in (8). The global result states that in the presence of AO's, the RLM_1 statistic retains its asymptotic χ^2 distribution, only multiplied with a constant of proportionality. This constant actually is a function of the autoregressive parameter ϕ, the probability of occurrence of AO's π and the (absolute) magnitude of the outliers ζ. The second result is called local, because it describes the behavior of the RLM_1 test for $\pi \downarrow 0$, i.e., for (infinitesimally) small fractions of outliers. In this way, the result can be compared to the derivation of an influence curve, see Hampel *et al.* (1986) for influence functions of estimators in the regression context, Martin and Yohai (1986) for influence functions of estimators in the time series context, and Peracchi (1991) for influence functions of test statistics in the regression context. The present result complements the results in these articles by presenting the influence of infinitesimally small fractions of contamination on the distribution of a test statistic in the time series context. A proof of the theorem is given in the appendix.

Theorem 1 *Consider the AR(1) model without a constant under symmetric additive outlier contamination and standard Gaussian innovations. If ψ is odd and $w_y(\cdot) \equiv 1$, then*

$$\lim_{T\to\infty} RLM_1 \xrightarrow{d} c_G \cdot \chi_1^2 \ ,$$

where χ_1^2 denotes a random variate with a chi-squared distribution with one degree of freedom, and where c_G is equal to the ratio of (25) and (26).

Furthermore, for π sufficiently small,

$$\lim_{T\to\infty} RLM_1 \xrightarrow{d} (1 + c_L \cdot \pi + O(\pi^2)) \cdot \chi_1^2 \,,$$

and where c_L is a function of ζ that is equal to the sum of the right-hand sides of (34) and (35), respectively.

Corollary 1 *For the OLS-based LM_1 test, the value of c_L in Theorem 1 reduces to*

$$\begin{aligned} c_L \;=\; & \frac{1}{3\sigma_x^4}\left(\phi^2\zeta^6 + (1+6\phi^2\sigma_x^2)\zeta^4 + (3\phi^2\sigma_x^4 + 6\sigma_x^2 + 3\sigma_x^4)\zeta^2\right) + \\ & \frac{-1}{3\sigma_x^4}\zeta^2(\zeta^2 + (7+\phi^2)\sigma_x^2) + \\ & \frac{2\phi\zeta^2}{3}(6\phi^2 - \zeta^2 - 3\phi^2\sigma_x^2), \end{aligned} \tag{19}$$

with $\sigma_x^2 = (1-\phi^2)^{-1}$. As a result, $c_L = (1-\phi^2)^2\phi^2\zeta^6/3 + O(\zeta^4)$.

Remark 1: Theorem 1 treats the case of M estimators. If GM estimators are used, the qualitative result remains the same, only now the influence curve of the scale estimator of y_{t-1} enters into the formulas. This illustrates that this function must be bounded in order for the influence function of the test to be bounded. Note that the influence function of the location of y_{t-1} does not enter, as this function is identically equal to zero because of the symmetry of the additive outliers.

Remark 2: The type of standard errors that is used in the computation of the LM_1 test is crucial for the local robustness properties of the test. If we use ordinary standard errors such as in Theorem 1, the constant c_L for the OLS-based test is an unbounded function of ζ. This function increases to plus infinity for $\zeta \to \pm\infty$. In other words, for local contamination the OLS-based test has a level above the nominal level. In contrast, we can use heteroskedasticity consistent (HCC) standard errors as in White (1980). This amounts to replacing the specification of k_2 in the proof of Theorem 1 by $k_2 = E(y_{t-1}^4\psi(\varepsilon_t)^2)$. Consequently, (27) in the appendix reduces to 1 and is independent of π. Therefore, the first two lines in the expression for c_L in (19) vanish and the dominant term in the expression for c_L for the OLS-based test becomes $-2\phi\zeta^4/3$. This is an unbounded function of ζ that tends to minus infinity for $\zeta \to \pm\infty$. In other words, the OLS-based test with HCC standard errors has a level below the nominal level under local contamination. Also, the order of c_L as a function of ζ is smaller if HCC standard errors are used, which signifies that the use of such standard errors alleviates part of the nonrobustness of the OLS-based LM_1 test.

For the OLS-based test, a closed expression for the constant of proportionality c_G is available, but too lengthy to present here. This constant actually is a function of the other parameters of the model, $c_G = c_G(\phi, \zeta, \pi)$, which satisfies (i) $c_G(\phi, 0, \pi) = 1$ for all combinations of ϕ and π, (ii) $\partial c_G/\partial\zeta > 0$ for $\zeta \in (0, q(\phi, \pi)]$, with q some function of ϕ and π, (iii) $\partial c_G/\partial\zeta < 0$ for $\zeta \in (q(\phi, \pi), \infty)$ and (iv) $c_G(\phi, \zeta, \pi) \to 1$ as $\zeta \to \infty$. These properties of the function $c_G(\phi, \zeta, \pi)$ show that the presence of AO's makes the LM_1 test biased towards the alternative, i.e. one finds spurious nonlinearity when neglecting AO's. Only if the size of the outliers becomes very large, the asymptotic distribution returns to the standard $\chi^2(1)$ distribution. Figure 2 graphs c_G for $\phi \in [0, 0.9]$, $\zeta \in [0, 20]$ and $\pi = 0.01, 0.05, 0.10$ and 0.25. Apart from the properties stated above, it is seen that for these fractions of contamination, c_G increases as ϕ gets larger. Furthermore, for fixed ϕ and ζ, the constant is smaller for larger values of π. This suggests that as the fraction of contaminated observations increases, the distribution of the test statistic is distorted to a lesser extent. This, however, is not true. The null distribution of the test is less affected under large AO's, but the power of the test is severely distorted. This is easily seen by considering the nature of AO contamination. Given a time series x_t, we clutter the signal of this series using a white noise process $\zeta\delta_t$, see (2). If ζ is large, the white noise component in the contaminated series y_t dominates the signal of the original series x_t. Therefore, for large ζ the observed contaminated time series looks like the (imposed) white noise process, which results in a χ_1^2 distribution for the LM_1 test, irrespective of whether the original x_t process follows a STAR or an AR process. Put differently, under dominant AO contamination, the distribution of LM_1 collapses to the null distribution, even if the alternative holds true.

In practice, one usually includes a constant in the linear model under the null hypothesis. Its effect on the constant c_G is investigated by means of simulation. For several combinations of (ϕ, ζ, π), 1000 series of 100 observations are generated from a contaminated AR(1) model and for each series the LM_1 test statistic is computed. Next, the percentiles of the χ_1^2 distribution are regressed on the corresponding test outcomes to obtain an estimate of c_G. These estimates are shown in the upper graphs of Figure 3, where a bivariate kernel is used to obtain a relatively smooth surface. Comparing these graphs with Figure 2, the major difference seems to be that, for fixed ϕ and π, c_G attains its maximum for higher values of ζ, while the return to 1 after this point proceeds much slower. Apart from this, for $\pi = 0.05$, the maximum value of c_G is much higher when a constant is included.

The Monte Carlo experiment described above is repeated for the HBP-GM based test statistic, to obtain some insight in the behavior of the robust RLM_1 test. The estimates of c_G, set out in the lower graphs of Figure 3, are remarkably different from the corresponding estimates for the OLS-based test. It is easily seen that, at least for $\pi = 0.05$, the maximum value of c_G is much lower for the robust test, while the decrease following this maximum is much faster as well. In fact, for large values of ζ and ϕ, the constant drops below 1, resulting in the test statistic being slightly biased toward the null. Notably, for the robust test c_G is larger for $\pi = 0.10$ than for $\pi = 0.05$. This contrasts to the findings for the OLS-based test.

6 Monte Carlo experiments

Before applying our RLM_i tests to several empirical time series, we evaluate the performance of the tests by means of Monte Carlo simulations. We focus on two effects which arise due to robustifying the test statistics. First, we consider the effect of using a robust test in a setting where robustness is not required, i.e., in a setting without outliers. Second, we investigate the (relative) performance of the tests in the presence of additive outliers. In the Monte Carlo experiments, we restrict attention to the behavior of the $(R)LM_1$ test and to a probability of occurrence of outliers of 0.05. Results for the $(R)LM_2$ and $(R)LM_3$ statistics, as well as results for other contamination fractions are available on request from the corresponding author.

6.1 Monte Carlo design

In the Monte Carlo experiments, 1000 series of $T = 100$ observations of the core process x_t are generated either from the AR(1) model (4) or from a relatively simple LSTAR(1) model,

$$x_t = \phi x_{t-1} + (1 + \exp\{x_{t-1}\})^{-1}\theta x_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T, \tag{20}$$

where $\varepsilon_t \sim$ n.i.d.$(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon = 1$. The starting value x_0 is set equal to zero. In order to eliminate possible dependencies of the results on this initial condition, the first 100 observations are discarded. Contaminated series y_t are obtained by adding AO's to x_t according to the replacement model (1) with $\zeta_t = x_t + \zeta\delta_t$. Following Franses and Haldrup (1994), the variable δ_t takes the values -1, 0 and 1 with probability $\pi/2$, $1 - \pi$ and $\pi/2$, respectively. Thus, both positive and negative AO's occur with equal probability. The standard and robust nonlinearity tests are applied to both the clean and the contaminated series x_t and y_t, respectively.

6.2 Simulation results

In our experiments, we consider the effects of varying the autoregressive parameters in the AR and STAR models and the magnitude of the AO's. In practice, one has to decide on the order of the linear AR model assumed to hold under the null hypothesis. In our simulation study, we fix this order *a priori*. The upper panel of Table 1 shows rejection frequencies of the null hypothesis by the standard LM_1 and robust RLM_1 tests, using 5% critical values. The results in this part of the table are based on an AR(1) model with autoregressive parameter ϕ. AO's of magnitude $\zeta = 3$, 4 5 are added to the model with probability $\pi = 0.05$. In addition, the columns under $\zeta = 0$ show estimates of the size when the test is applied to the series without outliers.

It is seen for the clean series x_t that the rejection frequencies of both the standard and the robust test approximate the 5% signifance level quite well. However, in the presence of outliers, marked differences appear. For the standard test, the rejection frequencies increase to rather high levels for $\zeta = 5$ and large (absolute) values of ϕ. The distortions in the level of the robust test are much smaller: the rejection frequencies stay below 10%.

The bottom panel of Table 1 shows rejection frequencies for series generated by the LSTAR model (20) for various combinations of ϕ and θ. The probability and magnitudes of AO's are the same as above. Note that, since the transition function is symmetric around zero, ϕ and $\phi + \theta$ can be interchanged. The columns under $\zeta = 0$ again show estimates of the power of the tests applied to the core series. These figures show that the difference between the autoregressive parameters for the two extreme cases $f = 0$ and $f = 1$, i.e., θ, has to be considerable for the tests to be able to detect the nonlinearity. As expected, the power of the robust test is slightly lower than that of the nonrobust test, as the robust test downweights observations that are not outliers. The maximum difference, however, is below 10%, which is quite encouraging. The results when the tests are applied to the contaminated series y_t show that it pays off to use the robust test. Although the power of both the standard and the robust tests decreases for most choices of (ϕ, θ), the resulting drop is far larger for the standard tests (up to 40%). The robust version of the test performs better in 17 out of 21 cases.

The Monte Carlo results in this section suggest that the empirical performance of the tests is satisfactory. If applied to time series without outliers, the performance of the robust tests is similar to that of the more familiar OLS-based tests, with an expected slight power advantage for the OLS-based tests. Furthermore, in case of linear or nonlinear time series with not too many outliers, the robust tests point at the correct model more often than the standard test. Unreported results for other contamination fractions and the $(R)LM_2$ and $(R)LM_3$ tests concur with these findings.

7 Nonlinearity in industrial production

As stated in the introduction, regime-switching models have been applied in particular to study possible nonlinearity in business cycles. In this spirit Luukkonen and Teräsvirta (1991) , Teräsvirta and Anderson (1992) and Teräsvirta *et al.* (1994) consider modelling industrial production indices for a number of OECD countries by STAR models. In this section, we apply the standard and robust tests for nonlinearity to these series.

7.1 The sequence of tests

In general, specification of STAR models is done using the procedure of Teräsvirta (1994), see also Granger and Teräsvirta (1993) for an extensive discussion. This specification procedure consists of three stages. First, a linear AR model is specified to form the basis for further analysis. Secondly, the LM_2 test is carried out to test linearity against STAR. Although this test statistic was designed against a logistic alternative, Teräsvirta (1994) argues that it should have power against exponential STAR (ESTAR) models as well. To determine which variable(s) should be included in the transition function, one can consider special cases of the LM_2 test as follows. If the vector a in (14) is restricted to $a = (0, \ldots, 0, 1, 0, \ldots, 0)'$, i.e. $a_d = 1$ for some $d \in \{1, \ldots, p\}$ and $a_j = 0$ for all $j \neq d$, the vector q_t of auxiliary regressors in (15) only contains terms $\tilde{y}_t^{(p)} y_{t-d}$, $\tilde{y}_t^{(p)} y_{t-d}^2$ and $\tilde{y}_t^{(p)} y_{t-d}^3$. Thus, the auxiliary regression used to compute the simplified LM_2 statistic is

$$\hat{\varepsilon}_t = \phi' y_t^{(p)} + \theta_2' \tilde{y}_t^{(p)} y_{t-d} + \theta_3' \tilde{y}_t^{(p)} y_{t-d}^2 + \theta_4' \tilde{y}_t^{(p)} y_{t-d}^3 + u_t \, , \tag{21}$$

where $\hat{\varepsilon}_t$ are the residuals from the regression $y_t = \phi' y_t^{(p)} + \varepsilon_t, \theta_i = (\theta_{i1}, \ldots, \theta_{ip})'$ for $i = 2, \ldots, 4$. The null hypothesis to be tested is $H_0 : \theta_2 = \theta_3 = \theta_4 = 0$. The test is carried out for different values of d. If linearity is rejected for several values of d, the one with the smallest p-value is selected as transition variable. This rule is motivated by the notion that the test might be expected to have maximum power if the true transition variable is used.

The next step is to decide between the logistic and exponential transition functions. This can be done by a short sequence of tests nested within (21). The null hypotheses to be tested are $H_{04} : \theta_4 = 0$, $H_{03} : \theta_3 = 0 | \ \theta_4 = 0$, and $H_{02} : \theta_2 = 0 | \ \theta_3 = \theta_4 = 0$. If $\theta_4 = 0$, the model can only be an ESTAR model, see Teräsvirta (1994). Similarly, if $\theta_3 = 0$, the model can only be an LSTAR model. Granger and Teräsvirta (1993) suggest to carry out all three tests, independent of rejection or acceptance of the first or second test. The decision rule used to select the transition function then is: select an ESTAR model only if the p-value corresponding to H_{03} is smallest, choose an LSTAR model in all other cases.

7.2 Testing for nonlinearity in industrial production

The series studied by Luukkonen and Teräsvirta (1991), Teräsvirta and Anderson (1992) and Teräsvirta *et al.* (1994) are quarterly, seasonally unadjusted indices of industrial production for 11 OECD countries, covering the period 1960(i)-1986(iv). The data are made approximately stationary by taking seasonal differences of the logarithms.

First of all, we apply the LM_i tests, as discussed in section 4. The orders of the linear AR models under the null are taken from Teräsvirta and Anderson (1992), which have been selected using the Akaike Information Criterion. The p-values corresponding to the different tests are shown in Table 3.

Table 4 shows the results from applying both the standard and robust tests used to specify STAR models. Although the AR orders are taken from Teräsvirta and Anderson (1992), a number of differences are observed for the standard tests. For the series of France and Italy these arise due to corrections Teräsvirta and Anderson (1992) apply to correct for the effect of strikes that occurred in these countries. The sources of the different outcomes for Finland and Sweden are not clear.

Comparing the models which are selected by the standard and robust tests, it is seen that for the majority of countries different conclusions can be drawn. Only for Finland, Italy, Japan and the Netherlands exactly the same results are obtained. Inspection of the weights resulting from the HBP-GM estimator (not shown here) reveals that outliers seem to be present around the oil crises of 1973 and 1979 for all series, although the number and timing varies considerably across the different countries.

For Germany, Norway and Sweden, the results partially coincide. For Germany, the standard tests indicate a logistic model, while the robust tests seem to prefer an exponential model, although the p-values for the test sequence are quite close to each other (0.027, 0.016, and 0.023, respectively). Therefore, it would be sensible to estimate both types of models and compare them on, for example, forecasting ability. For Norway and Sweden, all tests select an LSTAR model, only the appropriate delay parameter differs.

The standard tests do not indicate the presence of STAR-type nonlinearity for the industrial production series of France, while the robust tests strongly suggest an ESTAR model, the p-values corresponding to H_{04}, H_{03}, and H_{02} are 0.227, 0.000, and 0.803, respectively.

Finally, for Austria, Belgium and the USA, the standard tests indicate that an LSTAR model might be appropriate for these series, whereas the robust tests are unable to reject linearity. Figures 4, 5, and 6 show the weights $w_r(r)$ that are assigned to the observations by the HBP-GM estimation procedure as well as the values of the transition function $f(y_{t-d}; \gamma, a, c)$ (taken from Teräsvirta and Anderson (1992) for Belgium and the USA and from Teräsvirta *et al.* (1994) for Austria). Inspection of these figures shows that the apparent nonlinearity is due to only a few outlying observations. Especially for Belgium and the USA, the observations which receive weight (close to) zero are located around the points where transitions from one regime to the other regime in the LSTAR model occur. These regime-shifts, which are seen to be very quick, are caused by the transition variable taking values opposite from the threshold c at consecutive points in time. Apparently, this coincides with y_{t-d} taking rather extreme values, which results in vertical outliers or bad leverage points, as suggested by the zero weights. For Belgium, these aberrant observations explain all regime-shifts, while for the USA the remaining ones which cannot be ascribed to outliers do not produce enough evidence for STAR-type nonlinearity.

8 Concluding remarks

In this paper we proposed robust LM-type tests for STAR nonlinearity. The tests, which are straightforward to compute, use a HBP-GM estimator to estimate the linear AR model under the null hypothesis. The Monte Carlo evidence suggests that the empirical performance of the tests is satisfactory. If applied to time series without outliers, they do not suffer from large size distortions or much loss of power. Furthermore, in case of linear or nonlinear time series with outliers, the robust tests point at the correct model more often than the standard tests. The application to a selection of the industrial production series indicates that one should carefully interpret evidence from standard tests, as the presence of only a few aberrant observations may cause spurious nonlinearity.

The results obtained in this paper point towards several directions for further research. The robust estimation techniques might be applied to construct robust tests for other types of nonlinearity. Motivated by the discussion on outliers and nonlinearity, an obvious possibility would be to consider robust testing for (G)ARCH. The clustering of large residuals typical in GARCH models might well be mimicked by an IO or a sequence of AO's. Alternatively, robust estimation of STAR models might be considered worthwhile, elaborating on results from Chan and Cheung (1994). These issues will be taken up in our subsequent research.

Appendix: proof of Theorem 1

Proof: We have the replacement model

$$x_t = \phi x_{t-1} + \eta_t, \tag{22}$$

$$y_t = x_t(1-\delta_t) + \zeta\delta_t, \tag{23}$$

where η_t is i.i.d. standard Gaussian distributed, $|\phi| < 1$, $x_0 = \eta_0/(1-\phi^2)^{1/2}$, and δ_t is i.i.d. with $P(\delta_t = 1) = P(\delta_t = -1) = \pi/2$, $P(\delta_t = 0) = 1-\pi$, and $0 \leq \pi \leq 1$. Furthermore, we have that

$$LM_1 = \frac{T^{-1}\left(\sum_{t=1}^{T} y_{t-1}^2\psi(\varepsilon_t)\right)^2}{\left(T^{-1}\sum_{t=1}^{T} y_{t-1}^4\right)\left(T^{-1}\sum_{t=1}^{T}\psi(\varepsilon_t)^2\right)}. \tag{24}$$

We first prove that the terms in the square in the numerator have expectation zero independent of the values of π, ζ, and ϕ. In order to see this, note that

$$E(y_{t-1}^2\psi(\varepsilon_t)) = E(y_{t-1}^2\psi(\eta_t + \zeta\delta_t - \tilde{\phi}\zeta\delta_{t-1} + (\phi-\tilde{\phi})x_{t-1})) = 0,$$

where $\tilde{\phi}$ denotes the GM estimator and where the last equality follows from the conditions stated in the theorem. Therefore, $T^{-1/2}\sum_{t=1}^{T} y_{t-1}^2\psi(\varepsilon_t)$ satisfies a central limit theorem and converges in distribution to a normal with mean zero and variance k_1, with

$$k_1 = E(y_{t-1}^4\psi(\varepsilon_t)^2) + 2\sum_{k=1}^{\infty} E(y_{t-1}^2 y_{t-k-1}^2\psi(\varepsilon_t)\psi(\varepsilon_{t-k})). \tag{25}$$

Define

$$k_2 = E(y_{t-1}^4)E(\psi(\varepsilon_t)^2), \tag{26}$$

then using the central limit theorem, LM_1 tends in distribution to

$$LM_1 \xrightarrow{d} \frac{k_1}{k_2}\chi_1^2.$$

Defining $c_G \equiv k_1/k_2$, this proves the first part of the theorem. Note that for $\pi = 0$ we have that $k_1 = k_2$. Therefore, the second part of the theorem is proved if we can show that $c_L = d(k_1/k_2)/dp$ evaluated in $\pi = 0$. This is similar to deriving the influence function of a statistic, see, e.g., Hampel et al. (1986).

In order to derive the local result, we split k_1/k_2 in two parts, namely

$$\frac{E(y_{t-1}^4\psi(\varepsilon_t)^2)}{k_2} \tag{27}$$

and

$$2\sum_{k=1}^{\infty}\frac{E(y_{t-1}^2 y_{t-k-1}^2\psi(\varepsilon_t)\psi(\varepsilon_{t-k}))}{k_2}. \tag{28}$$

In order to put together the result, we first present the necessary individual derivatives with respect to π. We make heavy use of the techniques for deriving influence functions in the time series context as presented in Martin and Yohai (1986). We obtain

$$\begin{aligned} \frac{d}{d\pi}E(\psi(\varepsilon_t^2))\bigg|_{\pi=0} &= E(\psi(\eta_t+\zeta)^2-\psi(\eta_t)^2)+ \\ &\quad E(\psi(\eta_t-\phi\zeta)^2-\psi(\eta_t)^2), \end{aligned} \tag{29}$$

$$\frac{d}{d\pi}E(y_{t-1}^4)\bigg|_{\pi=0} = \zeta^4+6\zeta^2E(x_{t-1}^2), \tag{30}$$

$$\begin{aligned} \frac{d}{d\pi}E(y_{t-1}^4\psi(\varepsilon_t)^2)\bigg|_{\pi=0} &= E((x_{t-1}+\zeta)^4\psi(\eta_t-\phi\zeta)^2-x_{t-1}^4\psi(\eta_t)^2)+ \\ &\quad E(x_{t-1}^4(\psi(\eta_t+\zeta)^2-\psi(\eta_t)^2)), \end{aligned} \tag{31}$$

$$\begin{aligned} \frac{d}{d\pi}E(y_{t-1}^2y_{t-2}^2\psi(\varepsilon_t)\psi(\varepsilon_{t-1}))\bigg|_{\pi=0} &= -E(x_{t-1}^3x_{t-2}^2\psi'(\eta_t)\psi(\eta_{t-1}))IF_{\tilde{\phi}}+ \\ & E((x_{t-1}+\zeta)^2x_{t-2}^2\psi(\eta_t-\phi\zeta)\psi(\eta_{t-1}+\zeta)), \end{aligned} \tag{32}$$

and

$$\frac{d}{d\pi}E(y_{t-1}^2y_{t-k-1}^2\psi(\varepsilon_t)\psi(\varepsilon_{t-k}))\bigg|_{\pi=0} = 0, \tag{33}$$

for $k > 1$, where

$$IF_{\tilde{\phi}} = \zeta E(\psi(\eta-\phi\zeta))/E(x_{t-1}^2\psi'(\eta_t))$$

is the (time series) influence function of the estimator for the autoregressive parameter under AO contamination and $\psi'(\cdot)$ is the first order derivative of $\psi(\cdot)$ with respect to its argument. The result now follows by observing that

$$\frac{d}{d\pi}\left(\frac{E(y_{t-1}^4\psi(\varepsilon_t)^2)}{k_2}\right)\bigg|_{\pi=0} = \frac{dE(y_{t-1}^4\psi(\varepsilon_t)^2)/d\pi - E(y_{t-1}^4)dE(\psi(\varepsilon_t)^2)/d\pi}{k_2} + \\ \frac{-E(\psi(\varepsilon_t)^2)dE(y_{t-1}^4)/d\pi}{k_2}\bigg|_{\pi=0} \tag{34}$$

and

$$\frac{d}{d\pi}\left(2\sum_{k=1}^{\infty}\frac{E(y_{t-1}^2 y_{t-k-1}^2\psi(\varepsilon_t)\psi(\varepsilon_{t-k}))}{k_2}\right)\bigg|_{\pi=0} = 2\frac{dE(y_{t-1}^2 y_{t-2}^2\psi(\varepsilon_t)\psi(\varepsilon_{t-1}))/d\pi}{k_2}. \tag{35}$$

□

References

Balke, N.S. and T.B. Fomby (1994) Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series, *Journal of Applied Econometrics*, **9**, 181–200.

Bustos, O.H. and V.J. Yohai (1986) Robust estimates for ARMA models, *Journal of the American Statistical Association*, **81**, 155–168.

Chan, W.S. and S.H. Cheung (1994) On robust estimation of threshold autoregressions, *Journal of Forecasting*, **13**, 37–49.

Chang, I., G.C. Tiao and C. Chen (1988) Estimation of time series parameters in the presence of outliers, *Technometrics*, **30**, 193–204.

Chen, C. and L-M. Liu (1993) Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association*, **88**, 284–297.

Coakley, C.W. and T.P. Hettmansperger (1993) A bounded influence, high breakdown, efficient regression estimator, *Journal of the American Statistical Association*, **88**, 872–880.

Davies, L. and U. Gather (1993) The identification of multiple outliers - with comments and rejoinder, *Journal of the American Statistical Association*, **88**, 782–801.

Davies, R.B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **64**, 247–254.

Davies, R.B. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **74**, 33–43.

Denby, L. and R.D. Martin (1979) Robust estimation of the first-order autoregressive parameter, *Journal of the American Statistical Association*, **74**, 140–146.

Franses, P.H. and N. Haldrup (1994) The effects of additive outliers on tests for unit roots and cointegration, *Journal of Business and Economic Statistics*, **12**, 471–478.

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

Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series subject to changes in regime, *Econometrica*, **57**, 357–384.

Hampel, H.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986) *Robust Statistics - The Approach based on Influence Functions*, John Wiley & Sons, New York.

Handschin, E., J. Kohlas, A. Fiechter and F. Schweppe (1975) Bad data analysis for power system state estimation, *IEEE Transactions on Power Apparatus and Systems*, **2**, 329–337.

Hoek, H., A. Lucas and H.K. van Dijk (1995) Classical and Bayesian aspects of robust unit root inference, *Journal of Econometrics*, **69**, 27–59.

Huber, P.J. (1981) *Robust Statistics*, Wiley, New York.

Lucas, A., R. van Dijk and T. Kloek (1996) Outlier robust GMM estimation of leverage determinants in linear dynamic panel data models, mimeo, Tinbergen Institute.

Luukkonen, R. and T. Teräsvirta (1991) Testing linearity of economic time series against cyclical asymmetry, *Annales d'Économie et de Statistique*, **21**, 125–142.

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

Martin, R.D. (1981) Robust methods for time series, In D.F. Findley, editor, *Applied Time Series Analysis*, Academic Press, New York, pages 6843–759.

Martin, R.D. and V.J. Yohai (1986) Influence functionals for time series, *Annals of Statistics*, **14**, 781–818.

Peracchi, F. (1991) Robust m-tests, *Econometric Theory*, **7**, 69–84.

Rousseeuw, P.J. (1984) Least median of squares regression, *Journal of the American Statistical Association*, **79**, 871–880.

Rousseeuw, P.J. (1985) Multivariate estimation with high breakdown point, In W. Grossmann, G. Pflug, I. Vincze and W. Wertz, editors, *Mathematical Statistics and Applications, Vol. B*, Reidel Publishing, Dordrecht, pages 283–297.

Rousseeuw, P.J. and B.C. van Zomeren (1990) Unmasking multivariate outliers and leverage points - with comments and rejoinder, *Journal of the American Statistical Association*, **85**, 633–651.

Saikkonen, P. and R. Luukkonen. (1988) Lagrange multiplier tests for testing nonlinearities in time series models, *Scandinavian Journal of Statistics*, **15**, 55–68.

Simpson, D.G., D. Ruppert and R.J. Carroll (1992) On one-step GM estimates and stability of inferences in linear regression, *Journal of the American Statistical Association*, **87**, 439–450.

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

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

Teräsvirta, T., D. Tjøstheim and C.W.J. Granger (1994) Aspects of modelling nonlinear time series, In R.F. Engle and D.L. McFadden, editors, *Handbook of econometrics, vol. IV*, Elsevier Science, Amsterdam.

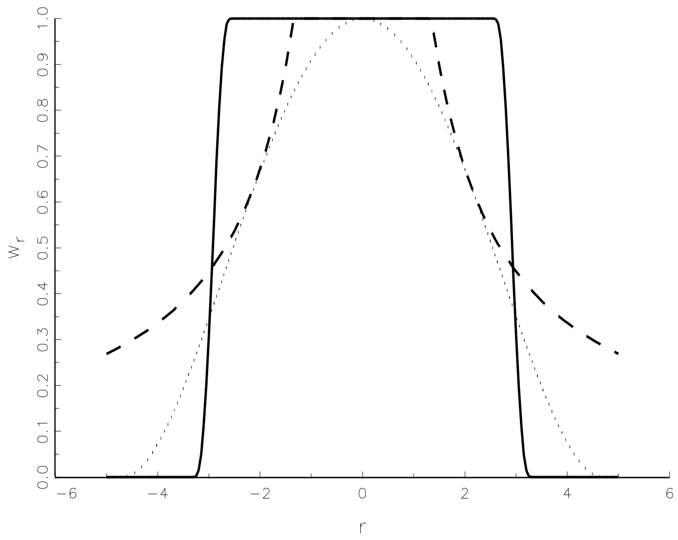
Tong, H. (1990) *Non-linear Time Series: a Dynamical Systems Approach*, Oxford University Press, Oxford.

Tsay, R.S. (1986*a*) Time series model specification in the presence of outliers, *Journal of the American Statistical Association*, **81**, 132–141.

Tsay, R.S. (1986*b*) Nonlinearity tests for time series, *Biometrika*, **73**, 461–466.

Tsay, R.S. (1989) Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association*, **84**, 231–240.

Figure 1: Weights for different ψ functions



Note: Weights $w_r(r) = \psi(r)/r$ resulting from (i) polynomial ψ function, $\psi(r) = r(1 - H(|r| - c_1))\mathrm{sgn}(r) + H(|r| - c_1)(1 - H(|r| - c_2)g(|r|)$, $c_1 = 2.576$, $c_2 = 3.291$ and $g(|r|)$ a fifth order polynomial such that $\psi(r)$ is twice continuously differentiable (solid line), (ii) Huber ψ function, $\psi(r) = \mathrm{med}(-c, c, r)$, $c = 1.345$ (dotted line) and (iii) Tukey's bisquare function, $\psi(r) = r(1 - (r/c)^2)^2 \cdot (1 - H(|r| - c)$, $c = 4.685$ (dashed line)

Table 1: Size and power of LM_1 and RLM_1 tests

		LM_1				RLM_1			
ϕ	$\phi+\theta$	$\zeta=0$	$\zeta=3$	$\zeta=4$	$\zeta=5$	$\zeta=0$	$\zeta=3$	$\zeta=4$	$\zeta=5$
-0.9	-0.9	4.9	11.4	17.6	23.2	6.1	8.7	9.3	7.8
-0.7	-0.7	4.3	10.0	15.6	19.7	5.0	6.6	8.5	9.3
-0.5	-0.5	2.9	8.2	10.8	12.8	3.8	5.7	7.6	8.5
-0.3	-0.3	3.4	5.8	6.5	7.7	3.8	5.8	6.3	5.9
-0.1	-0.1	3.7	4.8	4.9	4.9	4.1	4.5	5.5	4.7
0.0	0.0	4.1	5.1	4.2	4.4	3.9	4.6	5.3	5.2
0.1	0.1	4.2	5.6	4.5	4.4	4.2	4.9	5.4	5.3
0.3	0.3	3.6	6.5	6.6	6.7	3.9	6.1	6.0	6.4
0.5	0.5	3.6	8.0	10.7	11.8	4.4	7.4	9.5	7.8
0.7	0.7	2.7	10.0	16.2	20.2	4.2	8.8	10.2	10.4
0.9	0.9	1.5	10.6	19.0	26.0	3.4	8.9	9.3	7.2
-0.9	-0.7	16.2	16.2	18.8	21.8	14.5	12.3	13.9	15.2
-0.9	-0.5	32.7	19.4	20.1	20.8	29.7	19.2	24.3	26.7
-0.9	0.0	68.9	27.1	17.5	13.4	64.1	38.5	46.8	57.1
-0.9	0.5	90.2	40.7	22.0	13.4	85.5	59.9	69.6	79.4
-0.9	0.7	97.4	49.7	29.2	17.9	94.0	67.8	78.0	89.7
-0.9	0.9	98.3	66.5	46.2	29.6	96.6	79.6	81.9	87.7
-0.7	-0.5	9.9	12.3	15.4	17.6	8.7	9.8	13.2	14.1
-0.7	0.0	44.6	15.1	11.0	9.7	39.8	23.0	29.4	36.6
-0.7	0.5	79.7	30.0	16.0	9.3	73.9	45.8	54.1	69.4
-0.7	0.7	90.4	39.8	22.7	14.1	85.9	58.4	66.1	79.2
-0.7	0.9	96.3	55.2	36.4	26.5	93.3	71.3	74.8	82.2
-0.5	0.0	23.6	11.3	8.4	8.3	21.7	15.4	17.5	22.6
-0.5	0.5	64.6	25.0	14.1	8.4	58.8	36.7	48.7	54.7
-0.5	0.7	81.2	36.9	18.4	12.0	75.7	50.3	56.5	68.5
-0.5	0.9	92.3	54.1	34.3	23.6	87.1	62.0	65.3	74.3
0.0	0.5	20.6	9.9	8.4	7.5	18.4	12.5	16.8	18.9
0.0	0.7	42.8	17.9	12.6	10.5	39.1	24.1	30.5	33.8
0.0	0.9	62.5	38.0	30.1	26.0	60.9	44.3	45.3	49.4
0.5	0.7	7.0	12.0	16.1	17.9	7.5	12.6	14.9	11.5
0.5	0.9	18.7	21.6	23.9	24.9	20.9	23.3	25.1	21.7
0.7	0.9	6.2	17.1	24.2	27.5	8.4	15.2	16.4	14.9

Note: Rejection frequencies of LM_1 test (17) and RLM_1 test (18) at 5% significance level using F critical values for series generated by (20) with $\sigma_\epsilon^2 = 1$. Additive outliers are added with probability $\pi = 0.05$. The table is based on 1000 replications, $T = 100$.

Table 2: Size and power of LM_1 and RLM_1 tests

		LM_1				RLM_1			
ϕ	$\phi+\theta$	$\zeta=0$	$\zeta=3$	$\zeta=4$	$\zeta=5$	$\zeta=0$	$\zeta=3$	$\zeta=4$	$\zeta=5$
-0.9	-0.9	4.9	11.2	11.5	11.9	6.1	9.1	11.4	14.6
-0.7	-0.7	4.3	7.6	8.7	9.4	5.0	7.4	12.7	16.7
-0.5	-0.5	2.9	7.2	7.6	6.9	3.8	8.2	11.3	14.4
-0.3	-0.3	3.4	5.1	6.2	5.5	3.8	6.3	8.3	8.7
-0.1	-0.1	3.7	4.7	5.6	4.7	4.1	4.7	6.3	4.7
0.0	0.0	4.1	5.1	4.9	4.8	3.9	4.3	6.6	4.0
0.1	0.1	4.2	4.8	4.7	4.4	4.2	5.0	5.6	4.3
0.3	0.3	3.6	4.7	4.6	4.6	3.9	5.6	8.5	6.7
0.5	0.5	3.6	5.5	5.7	5.8	4.4	9.2	12.1	12.0
0.7	0.7	2.7	7.9	8.4	8.5	4.2	11.6	15.8	16.5
0.9	0.9	1.5	11.4	14.2	17.7	3.4	13.2	17.7	16.4
-0.9	-0.7	16.2	10.6	10.1	10.1	14.5	9.2	12.5	16.8
-0.9	-0.5	32.7	13.0	10.1	9.5	29.7	14.3	16.5	25.7
-0.9	0.0	68.9	14.1	8.5	6.6	64.1	20.5	32.6	46.5
-0.9	0.5	90.2	21.6	10.9	6.0	85.5	32.4	46.0	67.8
-0.9	0.7	97.4	29.0	14.4	9.5	94.0	39.7	50.2	74.6
-0.9	0.9	98.3	45.6	25.5	14.8	96.6	54.2	56.9	71.6
-0.7	-0.5	9.9	7.7	8.3	7.8	8.7	9.2	13.2	18.0
-0.7	0.0	44.6	8.4	5.7	5.6	39.8	13.3	20.8	30.7
-0.7	0.5	79.7	14.3	8.8	7.2	73.9	22.4	35.6	55.7
-0.7	0.7	90.4	21.0	12.0	8.4	85.9	30.6	41.0	65.7
-0.7	0.9	96.3	36.1	21.6	13.5	93.3	47.4	47.9	64.4
-0.5	0.0	23.6	7.8	7.1	6.4	21.7	8.5	14.8	19.0
-0.5	0.5	64.6	14.5	8.0	7.8	58.8	20.4	29.8	46.9
-0.5	0.7	81.2	18.6	9.4	6.6	75.7	30.6	36.9	53.9
-0.5	0.9	92.3	35.8	19.4	12.4	87.1	41.1	46.0	59.8
0.0	0.5	20.6	8.1	5.9	5.5	18.4	9.6	13.5	18.3
0.0	0.7	42.8	10.7	6.2	5.1	39.1	16.4	22.7	31.7
0.0	0.9	62.5	27.1	17.3	13.5	60.9	33.1	33.6	40.6
0.5	0.7	7.0	8.6	8.7	7.3	7.5	13.0	15.1	18.5
0.5	0.9	18.7	16.0	14.6	11.8	20.9	21.2	25.1	26.0
0.7	0.9	6.2	16.4	15.8	14.5	8.4	17.0	21.2	21.0

Note: Rejection frequencies of $LM_!$ test (17) and RLM_1 test (18) at 5% significance level using F critical values for series generated by (20) with $\sigma_\epsilon^2 = 1$. Additive outliers are added with probability $\pi = 0.10$. The table is based on 1000 replications, $T = 100$.

Table 3: p-values of LM-type tests for seasonal differences of quarterly industrial production data from 11 OECD countries 1961(i)-1986(iv)

		OLS			Robust		
Country	AR order	LM_1	LM_2	LM_3	RLM_1	RLM_2	RLM_3
Austria	5	0.056	0.209	0.082	0.845	0.341	0.801
Belgium	5	0.062	0.163	0.076	0.083	0.213	0.058
FR Germany	9	0.429	-	0.519	0.252	-	0.296
Finland	1	0.810	0.418	0.849	0.780	0.483	0.362
France	9	0.041	-	0.079	0.028	-	0.023
Italy	5	0.145	0.215	0.052	0.629	0.381	0.121
Japan	5	0.033	0.069	0.081	0.115	0.192	0.247
The Netherlands	1	0.209	0.123	0.128	0.954	0.994	0.991
Norway	8	0.363	-	0.521	0.076	-	0.150
Sweden	5	0.667	0.005	0.433	0.670	0.030	0.558
USA	6	0.014	-	0.072	0.191	-	0.228

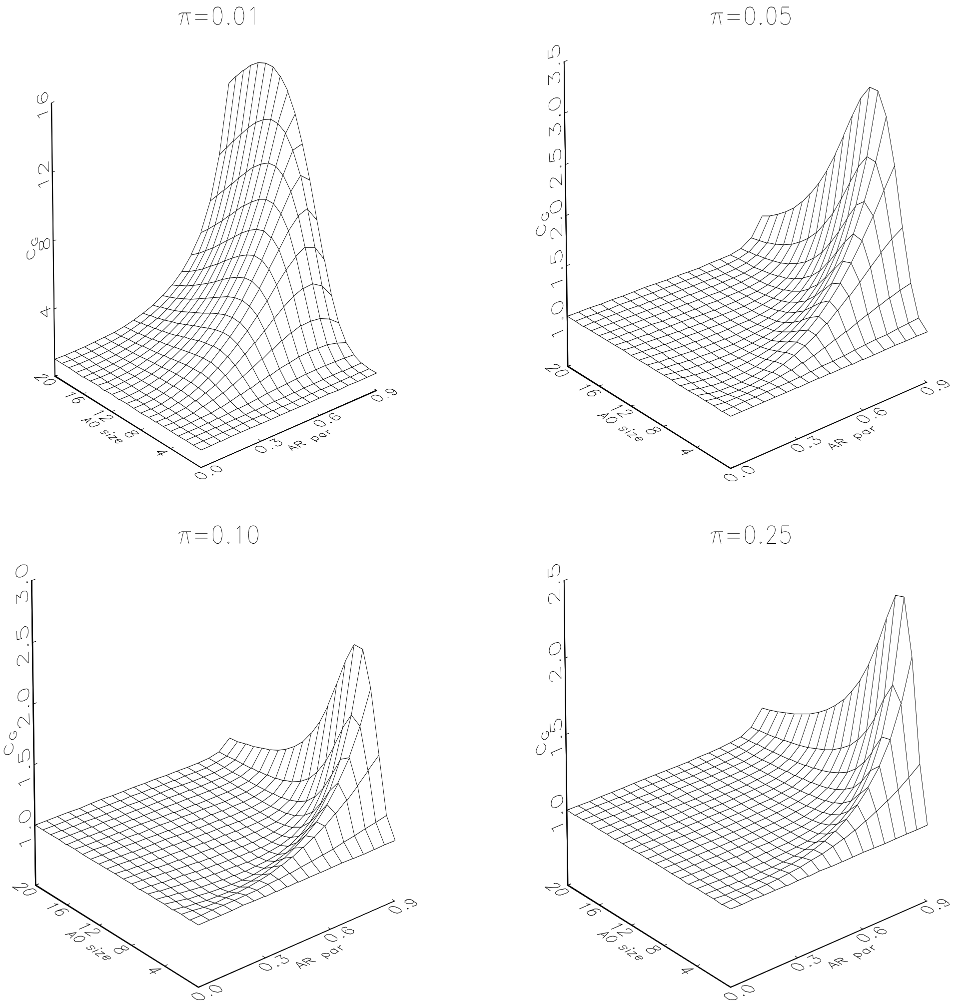
Note: AR orders have been taken from Teräsvirta and Anderson (1992). The standard and robust tests are F tests as discussed in sections 4 and 3. For Germany, France, Norway and USA the LM_2 test could not be computed due to multicollinearity problems.

Table 4: Model selection for seasonal differences of quarterly industrial production data from 11 OECD countries 1961(i)-1986(iv)

		Minimum p-value			Type of model(delay parameter)		
Country	AR order	TA	OLS	Robust	TA	OLS	Robust
Austria	5	0.010	0.010	0.260	LSTAR(1)	LSTAR(1)	Linear
Belgium	5	0.050	0.050	0.058	LSTAR(1)	LSTAR(1)	Linear
FR Germany	9	0.004	0.004	0.000	LSTAR(4)	LSTAR(4)	ESTAR(4)
Finland	1	0.547	0.393	0.376	Linear	Linear	Linear
France	9	0.156	0.099	0.007	Linear	Linear	ESTAR(2)
Italy	5	0.029	0.041	0.007	LSTAR(3)	LSTAR(1)	LSTAR(3)
Japan	5	0.000	0.000	0.038	L/ESTAR(1)	L/ESTAR(1)	LSTAR(1)
The Netherlands	1	0.123	0.123	0.153	Linear	Linear	Linear
Norway	8	0.031	0.031	0.012	LSTAR(5)	LSTAR(5)	LSTAR(3)
Sweden	5	0.016	0.011	0.050	LSTAR(3)	LSTAR(3)	LSTAR(4)
USA	6	0.006	0.006	0.061	LSTAR(3)	ESTAR(3)	Linear

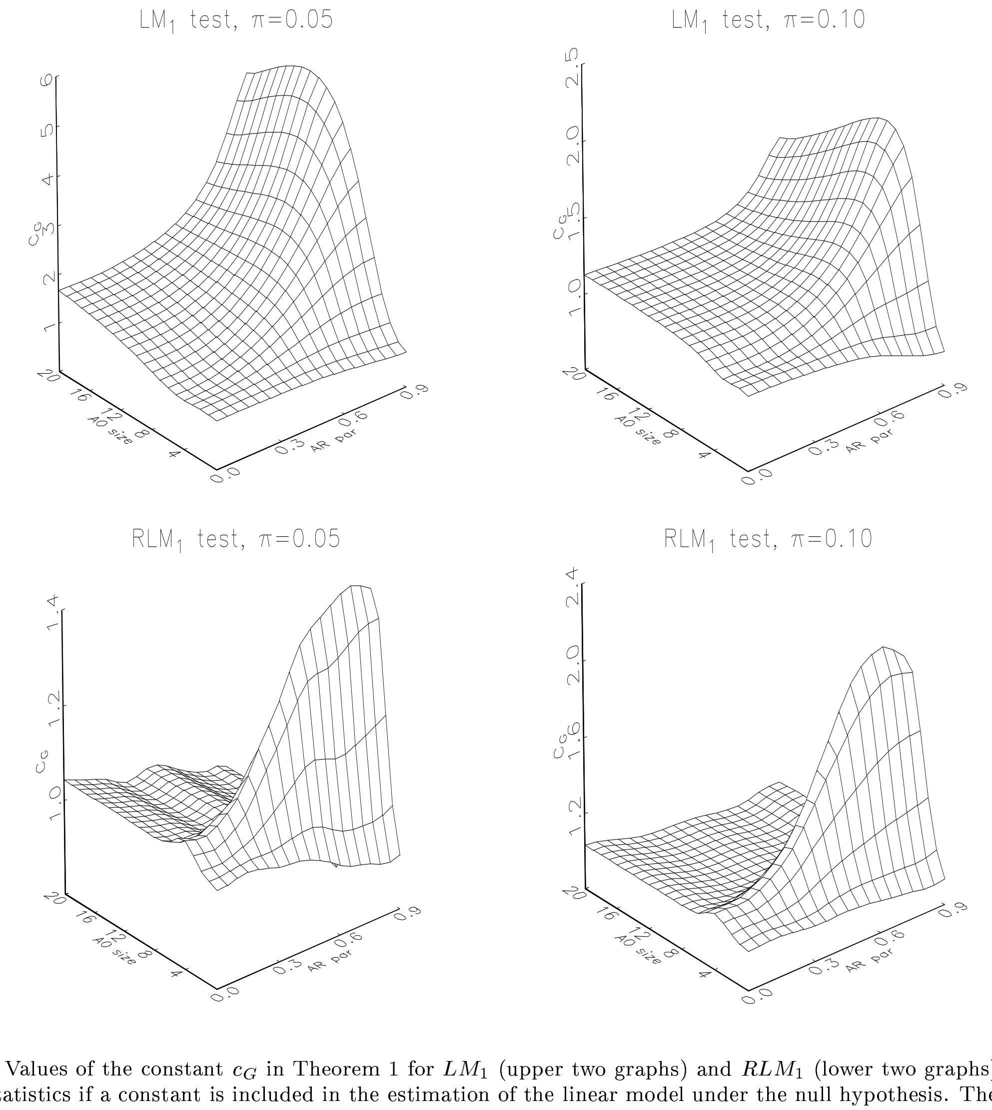
Note: AR orders have been taken from Teräsvirta and Anderson (1992). Their test results are given in the columns headed TA. The standard and robust tests are F tests based on (21). Minimum p-values are computed over $1 \leq d \leq 5$.

Figure 2: Constant c_G for LM_1 test



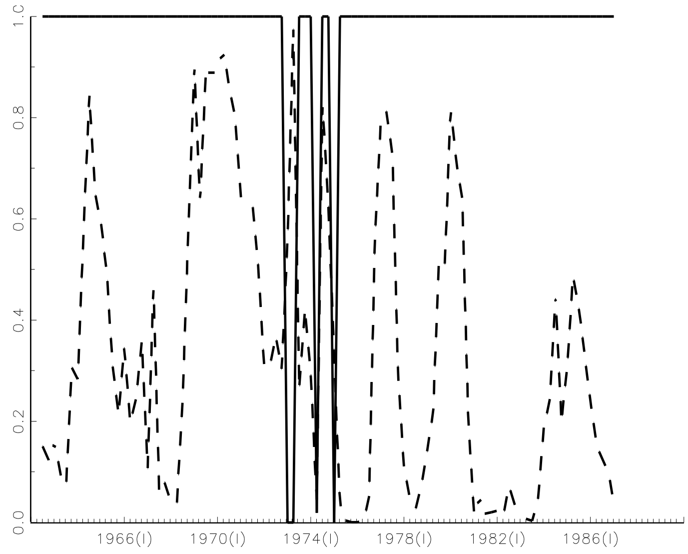
Note: Values of the constant c_G in Theorem 1 for the OLS-based LM_1 test statistic.

Figure 3: Constant c_G for LM_1 and RLM_1 tests, constant included



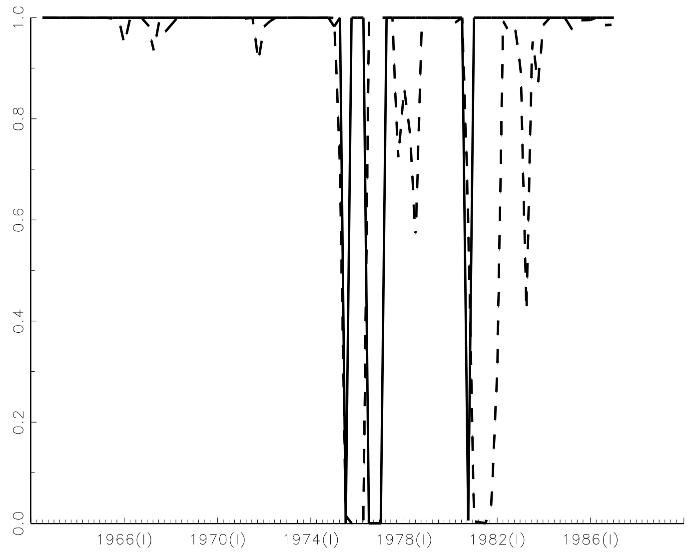
Note: Values of the constant c_G in Theorem 1 for LM_1 (upper two graphs) and RLM_1 (lower two graphs) test statistics if a constant is included in the estimation of the linear model under the null hypothesis. The figure is based on 1000 replications of a contaminated AR(1) model, $T = 100$.

Figure 4: Weights and transition function - Austria



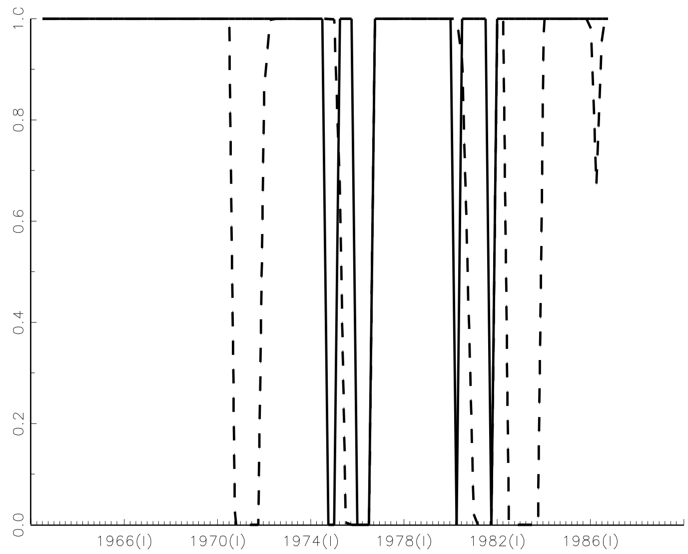
Note: Weights from robust estimation of AR(5) model used to compute nonlinearity tests (solid line) and transition function in fitted LSTAR model $f(x_{t-d};\gamma,a,c) = (1+\exp\{-\gamma(x_{t-d}-c)\})^{-1}$, $d = 1, \gamma = 2.2 \times 24, c = 0.063$ (dotted line)

Figure 5: Weights and transition function - Belgium



Note: Weights from robust estimation of AR(5) model used to compute nonlinearity tests (solid line) and transition function in fitted LSTAR model $f(x_{t-d};\gamma,a,c) = (1+\exp\{-\gamma(x_{t-d}-c)\})^{-1}$, $d = 1, \gamma = 7.3 \times 21.6, c = -0.015$ (dotted line)

Figure 6: Weights and transition function - United States



Note: Weights from robust estimation of AR(6) model used to compute nonlinearity tests (solid line) and transition function in fitted LSTAR model $f(x_{t-d}; \gamma, a, c) = (1 + \exp\{-\gamma(x_{t-d} - c)\})^{-1}, \quad d = 3, \gamma = 49 \times 17.5, c = 0.0061$ (dotted line)