DÉPARTEMENT DE SCIENCE ÉCONOMIQUE
DEPARTMENT OF ECONOMICS

CAHIERS DE RECHERCHE / WORKING PAPERS

0005E

Searching for Additive Outliers in Nonstationary Times Series

by
Pierre Perron and Gabriel Rodriguez

ISSN: 0225-3860

uOttawa

Faculté des sciences sociales
Faculty of Social Sciences

CP 450 SUCC. A
OTTAWA (ONTARIO)
CANADA K1N 6N5

P.O. BOX 450 STN. A
OTTAWA, ONTARIO
CANADA K1N 6N5

Searching for Additive Outliers In Nonstationary Time Series*

Pierre Perron — Boston University

Gabriel Rodríguez — Université d'Ottawa

May 9, 2000; Revised: August 25, 2001

Abstract

Recently, Vogelsang (1999) proposed a method to detect outliers which explicitly imposes the null hypothesis of a unit root. It works in an iterative fashion to select multiple outliers in a given series. We show, via simulations, that under the null hypothesis of no outliers, it has the right size in finite samples to detect a single outlier but when applied in an iterative fashion to select multiple outliers, it exhibits severe size distortions towards finding an excessive number of outliers. We show that his iterative method is incorrect and derive the appropriate limiting distribution of the test at each step of the search. Whether corrected or not, we also show that the outliers need to be very large for the method to have any decent power. We propose an alternative method based on first-differenced data that has considerably more power. We also show that our method to identify outliers leads to unit root tests with more accurate finite sample size and robustness to departures from a unit root. The issues are illustrated using two US/Finland real-exchange rate series.

Keywords: Additive Outliers, t-test, Wiener process, unit root, size, power.

JEL: C2, C3, C5

*This paper is drawn from chapter 3 of Gabriel Rodríguez's PhD dissertation at the Université de Montréal, Rodríguez (1999). We would like to thank Tim Vogelsang for useful conversations. We also thank Lynda Khalaf for comments on an earlier version of this paper entitled "Additive Outliers and Unit Roots with an Application to Latin-American Inflation" when it was presented at the 39th Congrès de la société canadienne de sciences économiques, Hull (Québec), May 1999.

Address for correspondence: Pierre Perron, Department of Economics, Boston University, 270 Bay State Road, Boston, MA, 02215, USA (e-mail: perron@bu.edu).

1 Introduction

From Fox (1972), who introduced the notion of additive and innovational outliers, issues related to this type of atypical observations in time series have received considerable attention in the statistics and econometric literature. The outlier detection issue, itself, has received particular attention [1]. Another topic of interest in the research has been the estimation of $ARMA$ models in the presence of outliers. In this case, as mentioned by Chen and Liu (1993), a common approach is to identify the locations and the types of outliers and then to accommodate the effects of outliers using intervention models as proposed by Box and Tiao (1975). This approach requires iterations between stages of outlier detection and estimation of the model [2].

In the context of integrated data (processes with an autoregressive unit root), the effects of additive outliers have recently been the object of sustained research. It is by now well recognized that outliers affect the properties of unit root tests (e.g., Franses and Haldrup (1994)). They do so by inducing a negative moving average component in the noise function which causes most unit root tests to exhibit substantial size distortions towards rejecting the null hypothesis too often. Franses and Haldrup (1994) suggested applying Dickey-Fuller (1979) unit root tests by incorporating dummy variables in the autoregression chosen on the basis of the outlier detection procedure proposed by Chen and Liu (1993). This procedure has been implemented in the computer program $TRAM$ (Time Series Regression with $ARIMA$ Noise and Missing Values) written by Gómez and Maravall (1992b), which allows us to estimate $ARIMA$ models where missing observations may be treated as additive outliers.

In an interesting recent paper, Vogelsang (1999) makes two contributions to the issue about the effects of additive outliers on unit root tests. First, recognizing that outliers induce a negative moving average component, he suggests using unit root test developed by Stock (1999) and Perron and Ng (1996) that are robust, in terms of achieving exact size close to nominal size in small samples, even in the presence of a substantial negative moving average component. He shows via simulations that these unit root tests are little

[1] See, e.g., Hawkins (1980) who presents a set of methods proposed before 1980 and Hawkins (1973) who proposed one of the most used methods, based on order statistics, to detect for outliers.

[2] Some references are Chang, Tiao and Chen (1988) and Tsay (1986). Chen and Liu (1993) also followed this way and they proposed another method to detect the locations of the outliers and the joint estimation of the parameters of the model. Their point of view was the fact that even if the model is well specified, outliers may still produce biased estimates of the parameters and, hence, may affect the outlier detection procedure. This is because atypical observations, in general, affect the variance of the estimates (e.g., Peña (1990)).

affected by systematic outliers. Secondly, he recognized that one can take advantage of the null hypothesis of a unit root in devising an outlier detection procedure. This allows the derivation of a non-degenerate limiting distribution for the t-statistic on the relevant one-time dummy.

In this paper, we make further contributions following the second suggestion of Vogelsang (1999). We show, via simulations, that Vogelsang's (1999) procedure, under the null hypothesis of no outlier, has the right size in finite samples to detect a single outlier but, when applied in an iterative fashion to select multiple outliers, it exhibits severe size distortions towards finding an excessive number of outliers. We show that there is a basic flaw in the iterative method suggested by Vogelsang (1999). In effect, contrary to what he implicitly assumes, the limiting distribution of the test used is different at each iteration of the outlier detection procedure. We derive the appropriate limiting distribution and tabulate some critical values. When so corrected, his method is shown to have very low power to detect outliers (even a single one without the correction made) unless the magnitude of the outlier is very large. As an alternative, we propose a method based on first-differenced data which has considerably more power. All of the methods considered are illustrated using two US/Finland real-exchange rate series.

The rest of the paper is organized as follows. Section 2 deals with the model and the issue of outlier detection. It reviews the procedure suggested by Vogelsang (1999) and presents simulation evidence about its size. Section 3 derives the correct limiting distribution of the test he suggested for each iteration of the outlier detection procedure. Section 4 presents the procedure based on first-differenced data. Section 5 compares its size and power to methods based on levels of the data using simulations. The size of unit root tests corrected using the various methods to detect outliers is investigated in Section 6. An empirical illustration using two US/Finland real-exchange rate series is presented in Section 7. Section 8 presents brief concluding remarks and some details about the data used are discussed in an appendix.

2 The model and the issue of outlier detection

There is a large literature in statistics and econometrics on the subject of outlier detection in $ARMA$ models. The standard approach is to estimate a fully parameterized $ARMA$ model and construct a t-statistic for the presence of an outlier. Such a t-statistic is constructed at all possible dates and the supremum is taken. The value of the supremum is then compared to a critical value to decide if an outlier is present. Some references are Tsay (1986), Chang, Tiao and Chen (1988), Shin, Sharkar and Lee (1996) and Chen and Liu (1993). Using a time series with an $ARIMA$ noise function, Gómez and Maravall (1992a) proposed to analyze missing observations as additive outliers. This paper was the basis for the computer program TRAM written by Gómez and Maravall (1992b) to estimate $ARIMA$ models with missing observations, which was used by Franses and Haldrup (1994) in the context of outlier detection in time series with unit roots.

The issue of outlier detection in the unit root framework offers a distinct advantage, namely that one can work under the null hypothesis that a unit root is present. This is the approach taken by Vogelsang (1999) whose procedure has two useful features. First, it does not require a fully parametric model of the noise function and is valid for a wide class of processes. Second, an asymptotic distribution can be obtained and critical values tabulated even without having to make specific distributional and parametric assumptions about the data-generating process.

The data-generating process entertained is of the following general form:

$$y_t = d_t + \sum_{j=1}^{m} \delta_j D(T_{ao,j})_t + u_t \tag{1}$$

where $D(T_{ao,j})_t = 1$ if $t = T_{ao,j}$ and 0 otherwise. This permits the presence of m additive outliers occurring at dates $T_{ao,j}$ $(j = 1, ..., m)$. The term d_t specifies the deterministic components. In most cases, $d_t = \mu$ if the series is non-trending or $d_t = \mu + \beta t$ if the series is trending (of course, other specifications are possible). The noise function is integrated of order one, i.e.

$$u_t = u_{t-1} + v_t \tag{2}$$

where v_t can be, for example, a linear process of the form $v_t = \varphi(L)e_t$ with $\varphi(L) = \sum_{i=0}^{\infty} \varphi_i L^i$ $(\sum_{i=0}^{\infty} i^2 \varphi_i^2 < \infty)$ and e_t is a martingale difference sequence with mean 0 and $\sigma_e^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(e_t^2)$ is finite. What is important is that the sequence v_t satisfies the condition for the application of a functional central limit theorem such that $T^{-1/2} \sum_{t=1}^{[Tr]} v_t \Rightarrow \sigma W(r)$ where $W(r)$ is the unit Wiener process, $\Rightarrow$ denotes weak convergence in distribution and $\sigma^2 = \lim_{T\to\infty} T^{-1} E(\sum_{t=1}^{T} v_t)^2$ with $0 < \sigma^2 < \infty$.

The detection procedure, suggested by Vogelsang (1999), starts with the following regression estimated by OLS (if necessary, a time trend can also be included),

$$y_t = \widehat{\mu} + \widehat{\delta} D(T_{ao})_t + \widehat{u}_t \tag{3}$$

where $D(T_{ao})_t = 1$ if $t = T_{ao}$ and 0 otherwise. Let $t_{\widehat{\delta}}(T_{ao})$ denote the t-statistic for testing $\delta = 0$ in (3). Following Chen and Liu (1993), the presence of an additive outlier can be tested using

$$\tau = \sup_{T_{ao}} \mid t_{\widehat{\delta}}(T_{ao}) \mid .$$

Assuming that $\lambda = T_{ao}/T$ remains fixed as T grows, Vogelsang (1999) showed that as $T \to \infty$,

$$t_{\widehat{\delta}}(T_{ao}) \Rightarrow H(\lambda) = \frac{W^*(\lambda)}{(\int_0^1 W^*(r)^2 dr)^{1/2}} \tag{4}$$

where $W^*(\lambda)$ denotes a demeaned standard Wiener process (i.e. $W^*(\lambda) = W(\lambda) - \int_0^1 W(s)ds$). If (3) also includes a time trend, $W^*(\lambda)$ will denote a detrended Wiener process. Furthermore, from the continuous mapping theorem it follows that,

$$\tau \Rightarrow \sup_{\lambda \in (0,1)} |H(\lambda)| \equiv H^*. \tag{5}$$

The distribution given in (5) is non-standard but is invariant with respect to any nuisance parameters, including the correlation structure of the noise function. The asymptotic critical values for τ were obtained using simulations. The Wiener processes were approximated by normalized sums of *i.i.d.* $N(0,1)$ random deviates using 1000 steps and 50,000 replications. Two cases were considered according to the deterministic components included in (3). When there is an intercept in (3) the critical values are 3.53, 3.11 and 2.92 at the 1, 5 and 10% significance levels, respectively. If a time trend is also included in (3) the corresponding critical values are 3.73, 3.31 and 3.12.[3]

The outlier detection procedure recommended by Vogelsang (1999) is implemented as follows[4]. First, compute the τ statistic for the entire series and compare τ to the appropriate critical value. If τ exceeds the critical value, then an outlier is detected at date $\widehat{T}_{ao} = \arg\max_{T_{ao}} |t_{\widehat{\delta}}(T_{ao})|$. The outlier and the corresponding row of the regression is dropped and (3) is again estimated and tested for the presence of another outlier. This continues until the test shows a non-rejection.

2.1 Simulation experiments for size

To asses the properties of the method in finite samples, we performed simulation experiments under the hypothesis that the series contain no outlier. We consider a simple data-generating process with an autoregressive unit root, i.e.

$$y_t = y_{t-1} + u_t.$$

[3]Critical values were also tabulated for the case where no deterministic components are included in (3). The critical values at 1%, 5% and 10% significance levels are 3.22, 2.84 and 2.65, respectively.

[4]This is equivalent to the *stepwise* procedure to select for multiple outliers. See Hawkins (1980).

Two cases are considered for the errors u_t; namely $MA(1)$ processes of the form $u_t = v_t + \theta v_{t-1}$ and $AR(1)$ processes of the form $u_t = \rho u_{t-1} + v_t$. In all cases, $v_t \sim i.i.d.\ N(0,1)$. We consider values of θ and ρ in the range $[-0.8, 0.8]$ with a step size of 0.2. Two sample sizes are used, $T = 100$ and $T = 200$. The number of replications used was $10,000$ and tests at the 5% and 10% significance levels were performed.

We first consider the size of the procedures in what we label the "one pass" case. The size is the number of times an observation is categorized as an outlier when searching for a single outlier (without iterating any further for a given sample). Results are presented in Table 1.

For the *i.i.d.* case, Vogelsang's method has an exact size close to nominal size. For the case with negative moving average errors, the test has size distortions (being liberal). These distortions are smaller when more deterministic components are included in the models. For positive moving average errors and particularly for the model that includes a time trend, the procedure is slightly undersized. A similar result is observed when there are positively correlated autoregressive errors.

The next experiments consider the properties of the method when applied in a full iterative fashion, i.e. continuing to search for additional outliers when one is found. Here, we record the total number of observations categorized as outliers divided by the number of replications. These values can be labelled as the expected number of outliers found. If the tests have the correct size α, say, at each steps of the iterations, and the tests are independent, this number should be close to $\alpha/(1-\alpha)$, that is .111 for a significance level 10% and .053 for a significance level 5%.

The results are presented in Table 2. The main thing to note is that Vogelsang's procedure finds many more outliers than would be expected if the test had the correct size at each steps. For example, for the model with only a constant with *i.i.d.* errors, $T = 100$, and a significance level of 10%, the number is .293 instead of .111, i.e. an average of 2.93 outliers for each replication which contains at least one outlier. These distortions increase when T increases to 200 with a value of .520 (instead of .111) which corresponds to approximately 5.2 outliers per replications which have at least one outlier.

3 The distribution of the test τ at different iterations

In the last section, we showed that the original procedure of Vogelsang (1999) has severe size distortions when applied in an iterative fashion to search for outliers. The reason for this is that the limiting distribution of the τ test given by (5) is only valid in the first step of the iteration. In subsequent steps, the asymptotic critical values used need to be modified. The correct limiting distribution at each step is given in the following Theorem.

Theorem 1 *Suppose that y_t is generated by (1) with $\delta_i = 0$ $(i = 1, ..., m)$ and let $\tau^{(i)}$ be the statistic τ obtained at step i of the iterative search for outliers, then*

$$\lim_{T\to\infty} \Pr[\tau^{(i)} > x] = \Pr[H^* > x]/\alpha^{i-1}$$

where α is the significance level of the test. Hence, the correct α-percentage point of the limiting distribution of $\tau^{(i)}$ is the α^i percentage point of the distribution of H^* defined by (5).

Proof: The basic reason for this result is that at different steps the tests are not independent; indeed they are asymptotically equivalent because of the fact that the series is integrated. Hence, at each step $\tau^{(i)} \Rightarrow H^*$ unconditionally on what happened in the previous steps. But subsequent steps are applied only if the previous one showed a rejection, hence one must consider the limiting distribution conditional upon a rejection at the previous step. For simplicity, consider this limiting distribution for the second step. It is given by, where x_α is the α-percentage point of the distribution of H^*,

$$\begin{aligned} \lim_{T\to\infty} \Pr[\tau^{(2)} &> x|\tau^{(1)} > x_\alpha] = \frac{\lim_{T\to\infty} \Pr[(\tau^{(2)} > x) \cap (\tau^{(1)} > x_\alpha)]}{\lim_{T\to\infty} \Pr[\tau^{(1)} > x_\alpha]} \\ &= \frac{\lim_{T\to\infty} \Pr[(\tau^{(2)} > x) \cap (\tau^{(1)} > x_\alpha)]}{\alpha} \end{aligned}$$

since $\tau^{(1)} \Rightarrow H^*$. Now, since we also have $\tau^{(2)} \Rightarrow H^*$,

$$\begin{aligned} \lim_{T\to\infty} \Pr[\tau^{(2)} &> x|\tau^{(1)} > x_\alpha] = \frac{\Pr[(H^* > x) \cap (H^* > x_\alpha)]}{\alpha} \\ &= \frac{\Pr[H^* > x]}{\alpha} \end{aligned}$$

provided $x \geq x_\alpha$, which we shall need to have tests with correct sizes. The result stated in the theorem follows using further iterations of the same arguments.■

We shall denote by τ_c the iterative outlier detection procedure that uses the correct (and different) asymptotic critical values at different steps. We have simulated some asymptotic critical values. We approximate the Wiener process by normalized sums of *i.i.d.* $N(0, 1)$ random variables using 200 steps. To obtain a fair range of critical values, we used 2 million replications. Nevertheless, even with such a large number of replications, the critical values can be obtained for only a few cases. This is because as we get further in the iterations of the outlier detection, we need percentage points of the distribution of H^* that are very far in the tail. For example, if the significance level is $\alpha = .05$, the percentage point needed at the 4th iteration is approximately .00001. Hence, even with 2 million replications we can only present critical values up to $i = 4$ for $\alpha = .05$, $i = 5$ for $\alpha = .10$, and $i = 7$ for $\alpha = .20$. These are presented in Table 3 [5].

4 A test using first differences of the data

As discussed in the next Section, Vogelsang's original procedure is not powerful unless the size of the outlier is very large. As a consequence, the full corrected iterative procedure is even less powerful since the critical values to be used at each iteration increase. Simulation evidence to that effect will be presented in the next section. Hence, it is desirable to entertain an alternative outlier detection procedure that is less likely to suffer from this low power problem.

We propose an iterative strategy using tests based on first-differences of the data. Consider data generated by (1) with $d_t = \mu$, and a single outlier occurring at date T_{ao} with magnitude δ. Then,

$$\Delta y_t = \delta[D(T_{ao})_t - D(T_{ao})_{t-1}] + v_t, \tag{6}$$

where $D(T_{ao})_t = 1$, if $t = T_{ao}$ (0, otherwise) and $D(T_{ao})_{t-1} = 1$, if $t = T_{ao} - 1$ (0, otherwise). If the data are trending a constant should be included. This reflects the fact that a unit root process with an outlier is characterized in first-differences by two successive outliers of equal magnitude but with opposite signs. We have that the least-squares estimate of δ is given by

$$\begin{aligned} \hat{\delta} &= (\Delta y_t - \Delta y_{t-1})/2 \\ &= (v_t - v_{t-1})/2 \end{aligned}$$

under the null hypothesis of no outlier. So the variance of $\hat{\delta}$ is given by

$$var(\hat{\delta}) = (R_v(0) - R_v(1))/2$$

where $R_v(j)$ is the autocovariance function of v_t at delay j. Let $\hat{R}_v(j) = T^{-1}\sum_{t=1}^{T-j} \hat{v}_t\hat{v}_{t+j}$ with $\hat{v}_t$ the least-squares residuals obtained from regression (6). Then, $\hat{R}_v(j)$ is a consistent estimate of $R_v(j)$. We can then consider the following test statistic

$$\tau_d = \sup_{T_{ao}} |t_{\hat{\delta}}(T_{ao})|$$

[5] Note that the critical values with $i = 1$ are not quite identical to those presented in Section 2 of this paper or in Vogelsang (1999) since 200 instead of 1000 steps where used to approximate the Weiner process. The differences, however, are minor and do not affect subsequent results.

where

$$t_{\hat{\delta}}(T_{ao}) = \hat{\delta}/((\hat{R}_v(0) - \hat{R}_v(1))/2)^{1/2}.$$

To detect multiple outliers, we can follow a strategy similar to that suggested by Vogelsang (1999), by dropping the observation labelled as an outlier before proceeding to the next step. The important feature is that, unlike for the case of tests based on levels (as the τ statistic of Vogelsang), in the limit the test τ_d is not perfectly correlated across each step of the iterations when dealing with multiple outliers. With *i.i.d.* errors, the values of τ_d are approximately uncorrelated at each steps of the iterations; with positively correlated errors, this no longer holds but the correlation is mild.

The disadvantage of this procedure, compared to that based on the level of the data, is that the limiting distribution depends on the specific distribution of the errors u_t, though not on the presence of serial correlation and heteroskedasticity. This problem is exactly the same as that for finding outliers in stationary time series since by differencing we effectively work with a stationary series. The standard practice in the literature is rather ad hoc and consists in rejecting if the t-statistic on some observation is greater than a critical value chosen to be some number between 3 and 4 (see, e.g., Tiao (1985), Chang and Tiao (1983) and Tsay (1986), among others). Here, we shall simulate critical values assuming *i.i.d.* normal errors and discuss the extent to which inference is affected when the data deviates from these specifications. So the data generating process is again

$$y_t = y_{t-1} + u_t \tag{7}$$

where $u_t \sim i.i.d.\ N(0,1)$. Two samples sizes are considered, namely $T = 100$ and $T = 200$. The number of replications used was 50,000. The percentage points of the test τ_d are presented in Table 4. To assess the size of the test in finite samples when correlation is present in the errors, we consider, as in Section 2.1, the same process defined by (7) with correlated errors. Two cases are considered for the errors u_t; namely $MA(1)$ processes of the form $u_t = v_t + \theta v_{t-1}$ and $AR(1)$ processes of the form $u_t = \rho u_{t-1} + v_t$. In all cases, $v_t \sim i.i.d.$ $N(0,1)$. We consider values of θ and ρ in the range $[-0.8, 0.8]$ with a step size of 0.4. The sample size is $T = 100$, the number of replications used was $10,000$ and tests at the 5% significance level were performed. We consider the iterative procedure with up to 4 outliers. The results are presented in Table 5.

The probability of finding at least one outlier is close to the nominal 5% level throughout. The test is slightly conservative with positive moving-average errors or when the autoregressive coefficient is very large in absolute value. The probability of finding at least two outliers is close to the theoretically expected value of .0025. The probability of finding more than 2 outliers is basically null in all cases. Hence, we conclude that the iterative procedure is adequate in that it delivers the expected number of rejections at each stage of the iterations. Also, the correction for the presence of serial correlation appears to perform satisfactorily.

5 Simulations for size and power

In this section, we present results about the size and, especially, the power of the various procedures when multiple outliers are present. The Data Generating Process considered is

$$y_t = \sum_{j=1}^{m} \delta_j D(T_{ao,j})_t + u_t, \tag{8}$$

$$u_t = u_{t-1} + v_t, \tag{9}$$

where $D(T_{ao,j}) = 1$ if $t = T_{ao,j}$ and 0 otherwise. Again, two cases are considered for the errors v_t; namely $MA(1)$ processes of the form $v_t = e_t + \theta e_{t-1}$ and $AR(1)$ processes of the form $v_t = \rho v_{t-1} + e_t$. In all cases, $e_t \sim i.i.d.\ N(0,1)$. We consider values of θ and ρ in the range $[-0.8, 0.8]$ with a step size of 0.4. This permits the presence of m additive outliers occurring at dates $T_{ao,j}$ $(j = 1, ..., m)$. We consider two cases, one with $m = 0$ to assess size and one with $m = 4$ outliers to assess power. All simulations are based on a sample size $T = 100$ and $10,000$ replications were performed. We present results only for the case where a constant is included in the set of deterministic components. The significance level of the test is set to 5%. For the procedures based on τ_c and τ_d, we used the critical values presented in Tables 3 and 4, respectively.

When $m = 4$, the location of the outliers are at observations $20, 40, 60,$ and 80. The magnitudes of the outliers considered are either a) $\delta_1 = 5$, $\delta_2 = 3$ and $\delta_3 = \delta_4 = 2$ or b) $\delta_1 = 10$ and $\delta_2 = \delta_3 = \delta_4 = 5$. We consider the properties of Vogelsang's uncorrected method (τ), its corrected version (τ_c) and the method based on first-differenced data (τ_d). The results are presented in Table 6 (MA errors) and Table 7 (AR errors).

Consider first the behavior of the tests when there is no outlier. The only procedure with a size close the expected theoretical nominal size (5% at the first step, .0025 at the second and basically 0 at the third and fourth) is that based on first-differenced data (τ_d), though as noted before it is somewhat conservative with an autoregressive coefficient that is large in absolute value and for positive MA coefficients. Vogelsang's procedure, whether corrected or not show substantial size distortion (liberal tests) in the presence of negative MA errors, and also to a lesser extent in the presence of strong negative AR errors. The results also confirm the fact that Vogelsang's uncorrected procedure (τ) finds an excessive number of outliers when applied in an iterative fashion.

The most interesting feature of the results is that the methods based on the level of the data have basically no power while the method based on first-differenced data has excellent power even for outliers of moderate size. Consider, for example, case (a) which is representative of outliers of moderate sizes. In the case with *i.i.d.* errors, Vogelsang's corrected procedure (τ_c) finds one outlier 14% of the cases while it basically never finds more than one outlier. The method based on first-differenced data finds at least one outlier almost 100% of the times and more than 3 outliers 23% of the times. Consider now case (b) which is representative of large outliers, the method τ_d finds 4 outliers basically 100% of the times, while the method τ_c finds at least one outlier 52% of the time and finds more than 2 outliers only 4% of the time. The results are qualitatively similar with errors that are serially correlated. Negative serial correlation (of the autoregressive or moving-average type) induces a loss of power while positive serial correlation (again of either type) induces an increase in power.

5.1 Robustness to departures from a unit root

It is of interest to asses the extent to which our suggested procedure is robust to departures from a unit root. Indeed, sometimes the purpose of detecting outliers is to provide appropriate corrections to unit root test, in which case the presence or not of a unit root is unknown. A popular device to analyze this issue is a so-called near-integrated process which specifies, under the null hypothesis of no outlier that

$$y_t = (1 + c/T)y_{t-1} + v_t \tag{10}$$

where v_t is a stationary process. Here, c is a non-centrality parameter which measures the extent of departures from a strict unit root process. When $c < 0$, we have a locally stationary process. This is labelled as a process local to unity, since as T increases the autoregressive parameter converges to one. A little algebra shows that, under this specification, the OLS estimate $\hat{\delta}$ from regression (6) is

$$\begin{aligned} \hat{\delta} &= (\Delta y_t - \Delta y_{t-1})/2 \\ &= [\frac{c}{T}(y_{t-2} + v_{t-1}) + (v_t - v_{t-1})]/2 \\ &= (v_t - v_{t-1})/2 + O_p(T^{-1/2}) \end{aligned}$$

Also, $(\hat{R}_v(0) - \hat{R}_v(1))/2$ remains a consistent estimate of the variance of $\hat{\delta}$. Hence, we can expect our procedure to remain adequate under this local to unit root setup. But the robustness of our procedure also extends to the case where y_t is a stationary processes. The reason is again that $(\hat{R}_v(0) - \hat{R}_v(1))/2$ remains a consistent estimate of the variance of $\hat{\delta}$.

To assess the size of the procedures τ_c and τ_d under departures from a unit root, we performed simulations from data generated by (10) with $T = 100$ and $v_t \sim i.i.d.\ N(0,1)$ for a range of values for c. The results are presented in Table 8. They clearly show that the procedure τ_d has an exact size close to the nominal 5% for any value of c. On the other hand the procedure τ_c shows increasing size distortions as c moves away from 0 (it is easy to show that the asymptotic distribution of the test τ_c is different under the local to unity framework; the Wiener process being replaced by an Orhnstein-Uhlenbeck process with drift parameter c). The last row of Table 8, shows the exact size when $y_t \sim i.i.d.\ N(0,1)$. Again, the procedure τ_d has an exact size close to the nominal 5%.

When the process is stationary, a more natural procedure is to base the test on a regression using levels of the data, i.e. a regression of the form

$$y_t = \mu + \delta D(T_{ao})_t + v_t. \tag{11}$$

Let δ^* be the OLS estimate of δ and denote the t-statistic for testing $\delta = 0$ by

$$t_{\delta^*}(T_{ao}) = \delta^*/\hat{R}_v(0)^{1/2}.$$

where $\hat{R}_v(0) = T^{-1}\sum_{t=1}^{T}\hat{v}_t^2$ with $\hat{v}_t$ the OLS residuals from regression (11). The statistic is then

$$\tau_l = \sup_{T_{ao}} |t_{\delta^*}(T_{ao})|.$$

To assess the size and power properties of τ_d and τ_l, we performed a simulation experiment with data generated by (8) with $\mu = 0$. Now the errors u_t are stationary and, again, two cases are considered; namely $MA(1)$ processes of the form $u_t = e_t + \theta e_{t-1}$ and $AR(1)$ processes of the form $u_t = \rho u_{t-1} + e_t$ with $e_t \sim i.i.d.\ N(0,1)$. The results, obtained from 10,000 replications, are presented in Tables 9 and 10 (the critical values for τ_l were obtained the same way as those for τ_d in Table 4 except that regression (11) is used instead of (6)). When the process is $i.i.d.$ or negatively serially correlated, the procedure τ_d based on first-differences is indeed less powerful than that based on level (τ_l). However, with positive serial correlation, the reverse holds and the procedure based on first-differences is more powerful. This is encouraging since most macroeconomic time series are positively correlated. Hence, for most applications of interest in economics, which have positive correlation, not only τ_d is valid if a unit root is present but is more powerful than the more common procedure based on level.

5.2 Consistency against local alternatives

It is well known that tests for outliers of the type considered here (based on the value of a single observation) are inconsistent against fixed alternatives. That is, the power of the tests against a fixed value of δ does not converge to one as the sample size increases. Nevertheless, insights into relative powers can be obtained looking at the properties of the tests in the presence of local alternatives of the form

$$H_1 : \delta_T = \delta_o T^a \tag{12}$$

for some fixed δ_o. It is easy to show that Vogelsang's procedure corrected or not (τ or τ_c) is consistent against alternatives of the form (12) only for values $a > 1/2$. On the other hand, the procedures τ_d and τ_l are consistent against such alternatives for any values $a > 0$. This goes some way towards explaining the greater power found for τ_d in the simulations.

5.3 Departures from normality

As we argued above, the procedure τ_d has several advantages over the procedure τ_c proposed by Vogelsang: the same critical values can be used at each step of the iterations, power is much higher, and it is robust to departures from a unit root. However, an advantage of the procedure τ_c is that is not affected by departures from the normality assumption (at least in large samples and with a strict unit root). This is not the case for the procedure τ_d whose distribution is heavily dependent on the normality assumption. This is not a new problem and it has been present throughout much of the literature on outlier detection and, as discussed in Section 4, most have resorted to recommend some ad hoc rule of thumbs to decide upon a rejection or not.

The distribution of tests like τ_d or the more common τ_l depends on the shape of the tail of the distribution of the error process. For alternative distributions, we obtained the following results from 25,000 replications. With uniform $[-1/2, 1/2]$ errors, which have no tail, the 5% critical value of τ_d is 2.61 for $T = 100$ (compared to 3.65 with Normal errors). On the other hand when the errors are distributed as a chi-square with one degree of freedom (centered to have mean zero), the 5% critical value is 6.22; indeed much higher due to the long right tail of the distribution.

To the authors' knowledge, no satisfactory procedure is available to overcome this dependence of outlier detection procedures on the exact nature of the error distribution. It may be possible to use extreme value theory and non-parametric estimates of tail behavior but such an extension is well beyond the scope of the present paper.

6 Size of corrected ADF unit root tests

As noted by Franses and Haldrup (1994) and Vogelsang (1999), the presence of outliers biases unit root tests towards over-rejection of the null hypothesis acting like a negative moving-average component. One of the aim of outlier detection mentioned by these authors is to be able to correct the unit root tests by incorporating appropriate dummy variables. To assess the relative merits of the outlier detection procedures discussed in correcting the size of unit root tests, we again resorted to a simulation analysis concerning the size of the Dickey-Fuller (1979) test given by the t-statistic for testing that $\alpha = 1$ in the following regression

$$y_t = \mu + \alpha y_{t-1} + \sum_{i=0}^{p+1}\sum_{j=1}^{m} \delta_{ij} D(T_{ao,j})_{t-i} + \sum_{i=1}^{k} d_i \Delta y_{t-i} + e_t.$$

where $D(T_{ao,j})_t = 1$ if $t = T_{ao,j}$ and 0 otherwise, with $T_{ao,j}$ $(j = 1, ..., m)$ the dates of the outliers identified. The data-generating process is the same as described earlier. In constructing the unit root tests, the lag length k was selected in the same way as in Vogelsang (1999), namely using a recursive general to specific t-test on the last lag with a significance level of 10% starting at some maximal value set at 5. The results are presented in Table 11 (MA errors) and Table 12 (AR errors). For each cases, the row marked "without" indicates the percentage of rejections of the null hypothesis that occurred when no outlier was selected; the row marked "with" indicates the percentage of rejections of the null hypothesis when outliers were detected and the appropriate dummies introduced in the autoregression; the row marked "total" is simply the sum of the two cases mentioned above.

We first consider the case where no outlier is present. This establishes a base case to compare size distortion with cases where outliers are present. With autoregressive errors all procedure have approximately the correct size. The same is true with a positively correlated moving-average component. As is well known the ADF unit root test suffers from substantial size distortion with a negatively correlated moving-average component and this is reflected in our results. When outliers are present, the size of the ADF test corrected for outliers using τ_c is, in almost all cases, larger than when corrected using the method τ_d. For example, with *i.i.d.* errors and large outliers, the size is .087 when corrected with τ_c and .041 when corrected with τ_d. Comparing the rows "with" and "without", we see that when outliers are present rejections of the unit root occurring when no outliers are identified are very small for the method τ_d while they are substantial when using the method τ_c.

As emphasized by Franses and Haldrup (1994), outliers induce an MA like component in the errors when they are not accounted for. Even if Vogelsang's method selects too few outliers (given its low power) the size of the corrected ADF test can be brought close to nominal size when using a data dependent method to select the lag length since the latter would tend to correct for missed outliers by choosing a higher lag length (since the missed outliers have the effect of inducing a negative moving-average structure in the errors). To verify this claim, we conducted the same simulation experiments with pure AR(1) errors and the lag length fixed at its true value 1. The results are presented in Table 13. Consider, for instance, the case with an autoregressive coefficient $\rho = -0.4$ with large outliers. The size of the unit root test corrected using τ_c is .19 with k fixed at 1 instead of .10 with k selected using the sequential t-test procedure. Hence, it is clear that the failure to account for all outliers present can be compensated by the selection of a larger lag length. Yet, as the results for the size of the unit root test corrected using τ_d show, a good method to select outliers does a better job at reducing size distortions.

7 Empirical applications

The procedures analyzed in the last sections were applied to two series of real-exchange rates for US/Finland. The first series covers the period 1900-1988 and it is constructed using the Consumption Price Index (CPI). The other series spans the years 1900-1987 and is constructed using the Gross Domestic Product (GDP) deflator. The series are shown in Figures 1 and 2, respectively. These are the same series used by Vogelsang (1999), Franses and Haldrup (1994) and Perron and Vogelsang (1992) and are described in more details in Appendix A.

Franses and Haldrup (1994) used the $TRAM$ program (Time Series Regression With $ARIMA$ Noise and Missing Values) written by Gómez and Maravall (1992b) to search for outliers in these two real-exchange rate series. They considered two types of outliers, additive outliers and outliers that produce temporary changes, denoted AO and TC outliers, respectively. For the US/Finland real-exchange rate series based on the CPI index, they found four additive outliers at dates 1918, 1922, 1945 and 1948. The observations associated with the years 1917, 1932 and 1949 were found to be outliers that produce temporary changes (TC outliers). For the US/Finland real-exchange rate series based on the GDP deflator, an additive outlier was found only at date 1918, whereas outliers that produce temporary changes were found at dates 1917, 1932, 1949 and 1957.

Table 14 reports the empirical results from applying the procedures discussed in this paper using 5% and 10% significance levels. Vogelsang (1999) presents results for additive outliers only for the US/Finland real-exchange rate series based on the CPI index. The dates he found (using the procedure τ) were 1917-1919, 1921 and 1932. When appropriately corrected, Vogelsang's (1999) method finds outliers only for the year 1918 at the 5% level and for 1918 and 1919 at the 10% level, illustrating the fact that when it is not corrected it tends to select more outliers than warranted. The procedure based on first-differenced data (τ_d) finds outliers at dates 1917, 1918, 1919, 1932 and 1948 at both the 5% and 10% significance levels. This illustrates how this latter method is more powerful.

For the US/Finland real-exchange rate series based on the GDP deflator, the method based on τ_c finds no outlier. As mentioned by Vogelsang (1999), this may be due to the presence of a shift in the mean of the series as documented by Perron and Vogelsang (1992). The procedure based on first-differenced data (τ_d) is, nevertheless, able to identify the years 1918 and 1948 as outliers at the 5% level [6]. These two dates are not associated with the change in mean identified by Perron and Vogelsang (1992) as occurring in 1937. The fact that our procedure identifies the year 1918 as an outlier is comforting since visual inspection clearly points in that direction.

8 Conclusions

We analyzed in this paper the size and power properties of some test procedures for multiple outliers in series with an autoregressive unit root. We showed, via simulations, that the procedure suggested by Vogelsang (1999) has indeed the right size when applied to detect a single outlier but that it finds an excessive number of outliers when applied in an iterative fashion. We showed this iterative method to be theoretically incorrect and we derived the appropriate limiting distribution for each step of the iterations. We also showed that, whether corrected or not such outlier detection methods based on the level of the data have very low power unless the magnitude of the outliers is unrealistically large. Our suggestion was to use a procedure based on first-differenced data which was shown to have considerably more power. Our analysis remained in the tradition of sequential searches for outliers. It may well be the case that a global procedure might perform better. Work is under way to investigate this issue.

[6] At the 10% significance level, the outliers found are for the years 1917, 1918, 1919, 1921, 1932, 1947, 1948 and 1957.

9 Appendix: The Data

The US/Finland real-exchange rate series based on the CPI index and the GDP deflator were kindly provided by Tim Vogelsang. They are the same series used in Vogelsang (1999), Franses and Haldrup (1994) and Perron and Vogelsang (1992). The US/Finland real-exchange rate series based on the CPI index is annual from 1900 to 1988, whereas that based on the GDP deflator is from 1900 to 1987. The details of the sources is as follows (see appendix A of Perron and Vogelsang (1992)): Nominal exchange rate series —1900-1988 from the Bank of Finland; CPI —1900-1985 from the Bank of Finland, 1986-1988 from the IMF (1988); GDP deflator —1900-1985 from the Bank of Finland, 1986-1987 from IMF (1988). The sources of the U.S. data are: for the GNP deflator —1869-1975 from Friedman and Schwartz (1982), 1976-1988 from IMF (1988); for the CPI —1860-1970 from the U.S. Bureau of the Census (1976) and 1971-1988 from IMF (1988).

References

[1] Box, G.E.P. and Tiao, G. C. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association* **70**, 70-79.

[2] Chang, I, and G. C. Tiao (1983), "Estimation of Time Series Parameters in the Presence of Outliers," Technical Report 8, University of Chicago, Statistics Research Center.

[3] Chang, I., Tiao, G. C. and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics* **30**, 193-204.

[4] Chen, C. and L. Liu (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association* **74**, 427-431.

[5] Dickey, D. A. and W. A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* **74**, 427-431.

[6] Fox, A. J. (1972), "Outliers in Time Series," *Journal of the Royal Statistical Association Series B* **43**, 350-363.

[7] Franses, P. H. and N. Haldrup (1994), "The Effects of Additive Outliers on Tests for Unit Roots and Cointegration," *Journal of Business & Economic Statistics* **12**, 471-478.

[8] Friedman, M. and Schwartz, A. J. (1982), *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices and Interest Rates, 1867-1975*, Chicago: The University of Chicago Press.

[9] Gómez, V. and A. Maravall (1992a), "Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman Filter," European University Institute, Working Paper ECO 92/80.

[10] Gómez, V. and A. Maravall (1992b), "Time Series Regression with ARIMA Noise and Missing Observations. Program TRAM," European University Institute, Working Paper ECO 92/81.

[11] Hawkins, D. M. (1973), "Repeated Testing for Outliers," *Statistica Neerlandica*, **27**, 1-10.

[12] Hawkins, D. M. (1980), *Identification of Outliers*, New York: Chapman and Hall.

[13] International Monetary Fund (1988), *International Financial Statistics: Yearbook*, Washington, DC.

[14] Peña, D. (1990), "Influential Observations in Time Series," *Journal of Business & Economic Statistics* **8**, 235-241.

[15] Perron, P. and S. Ng (1996), "Useful Modifications to Unit Root Tests with Dependent Errors and their Local Asymptotic Properties," *Review of Economic Studies* **63**, 435-463.

[16] Perron, P. and T. J. Vogelsang (1992), "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business & Economic Statistics* **10**, 301-320.

[17] Rodríguez, G. (1999), *Unit Root, Outliers and Cointegration Analysis with Macroeconomic Applications*, Unpublished PhD Dissertation, Département de Sciences Économiques, Université de Montréal.

[18] Shin, D. W., S. Sarkar and J. H. Lee (1996), "Unit Root Tests for Time Series with Outliers," *Statistics and Probability Letters* **30**, 189-197.

[19] Stock, J.H. (1999), "A Class of Tests for Integration and Cointegration," in Engle, R.F. and H. White (eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, pp. 137-167.

[20] Tiao, G. C. (1985), "Autoregressive Moving Average Models, Intervention Problems and Outlier Detection in Time Series," in E. J. Hannan, P. R. Krishnaiah, and M. M. Rao (eds.), *Time Series in the Time Domain, Handbook of Statistics 5*, New York: North Holland, 85-118.

[21] Tsay, R. S. (1986), "Time Series Model Specification in the Presence of Outliers," *Journal of the American Statistical Association* **81**, 132-141.

[22] U.S. Bureau of the Census (1976), *The Statistical History of the United States, From Colonial Times to the Present*, New York: Basic Books.

[23] Vogelsang, T. J. (1999), "Two Simple Procedures for Testing for a Unit Root when there are Additive Outliers," *Journal of Time Series Analysis* **20**, 237-252.

Table 1: Exact size of single outlier detection

MA Case						AR Case					
		Constant		Time trend				Constant		Time trend	
T	θ	5.0%	10.0%	5.0%	10.0%	T	ρ	5.0%	10.0%	5.0%	10.0%
100	-0.80	0.184	0.332	0.137	0.241	100	-0.80	0.073	0.149	0.068	0.135
	-0.60	0.101	0.191	0.095	0.177		-0.60	0.064	0.121	0.061	0.118
	-0.40	0.067	0.125	0.068	0.126		-0.40	0.053	0.107	0.055	0.106
	-0.20	0.050	0.104	0.051	0.097		-0.20	0.046	0.091	0.048	0.094
	0.00	0.043	0.082	0.043	0.082						
	0.20	0.038	0.076	0.037	0.075		0.20	0.031	0.073	0.036	0.071
	0.40	0.037	0.072	0.035	0.069		0.40	0.034	0.064	0.032	0.063
	0.60	0.037	0.069	0.035	0.067		0.60	0.023	0.050	0.029	0.053
	0.80	0.037	0.060	0.035	0.066		0.80	0.020	0.042	0.029	0.047
200	-0.80	0.227	0.400	0.177	0.327	200	-0.80	0.080	0.154	0.076	0.151
	-0.60	0.105	0.205	0.102	0.198		-0.60	0.064	0.126	0.065	0.124
	-0.40	0.065	0.131	0.068	0.127		-0.40	0.055	0.110	0.056	0.109
	-0.20	0.050	0.102	0.051	0.099		-0.20	0.048	0.097	0.049	0.096
	0.00	0.042	0.086	0.042	0.084						
	0.20	0.038	0.079	0.038	0.075		0.20	0.037	0.077	0.036	0.073
	0.40	0.037	0.076	0.036	0.072		0.40	0.034	0.067	0.030	0.062
	0.60	0.037	0.074	0.034	0.069		0.60	0.029	0.055	0.025	0.052
	0.80	0.037	0.074	0.034	0.069		0.80	0.022	0.043	0.022	0.041

Table 2: Expected number of outliers found using multiple outliers detection

MA Case						AR Case					
	Constant			Time trend				Constant		Time trend	
T	θ	5.0%	10.0%	5.0%	10.0%	T	ρ	5.0%	10.0%	5.0%	10.0%
100	-0.80	0.216	0.447	0.154	0.291	100	-0.80	0.124	0.278	0.091	0.198
	-0.60	0.139	0.306	0.109	0.220		-0.60	0.128	0.272	0.086	0.181
	-0.40	0.132	0.285	0.094	0.194		-0.40	0.127	0.286	0.088	0.192
	-0.20	0.129	0.292	0.092	0.196		-0.20	0.129	0.292	0.091	0.199
	0.00	0.129	0.293	0.096	0.205						
	0.20	0.127	0.295	0.097	0.208		0.20	0.128	0.294	0.097	0.204
	0.40	0.130	0.296	0.097	0.205		0.40	0.129	0.293	0.105	0.207
	0.60	0.133	0.288	0.101	0.207		0.60	0.136	0.294	0.110	0.227
	0.80	0.132	0.293	0.101	0.206		0.80	0.147	0.297	0.146	0.278
200	-0.80	0.295	0.638	0.206	0.428	200	-0.80	0.195	0.487	0.126	0.296
	-0.60	0.197	0.478	0.141	0.311		-0.60	0.203	0.499	0.133	0.301
	-0.40	0.200	0.494	0.138	0.308		-0.40	0.206	0.505	0.141	0.319
	-0.20	0.214	0.515	0.144	0.324		-0.20	0.217	0.519	0.143	0.328
	0.00	0.209	0.520	0.142	0.331						
	0.20	0.214	0.509	0.147	0.338		0.20	0.217	0.505	0.145	0.337
	0.40	0.212	0.505	0.147	0.341		0.40	0.217	0.494	0.145	0.332
	0.60	0.214	0.503	0.147	0.338		0.60	0.225	0.495	0.150	0.345
	0.80	0.216	0.500	0.148	0.339		0.80	0.221	0.488	0.183	0.359

Table 3: Asymptotic critical values of the test τ_c

α	i	Model 1 $z_t = \{1\}$	Model 2 $z_t = \{1, t\}$
0.05	1	2.99	3.33
	2	3.69	4.86
	3	4.29	13.16
	4	4.43	18.20
0.10	1	2.81	3.11
	2	3.38	3.94
	3	3.88	6.08
	4	4.33	14.43
	5	4.78	36.44
0.20	1	2.61	2.87
	2	3.05	3.41
	3	3.43	4.05
	4	3.79	5.40
	5	4.12	8.88
	6	4.42	18.04
	7	4.73	33.41

Table 4: Finite sample critical values of the test τ_d

Level of significance	Model 1 $z_t = \{1\}$		Model 2 $z_t = \{1, t\}$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$
1.0%	4.14	4.20	4.13	4.19
2.5%	3.87	3.95	3.85	3.94
5.0%	3.65	3.75	3.63	3.74
10.0%	3.44	3.56	3.42	3.55

Table 5: Exact Size of the test based on τ_d

		Probability to find			
		First outlier	Second outlier	Third outlier	Fourth outlier
i.i.d. Case		0.047	0.002	0.000	0.000
MA Case	$\theta = -0.80$	0.053	0.003	0.000	0.000
	$\theta = -0.40$	0.052	0.002	0.000	0.000
	$\theta = 0.40$	0.034	0.003	0.001	0.000
	$\theta = 0.80$	0.021	0.005	0.001	0.001
AR Case	$\rho = -0.80$	0.029	0.003	0.000	0.000
	$\rho = -0.40$	0.053	0.002	0.000	0.000
	$\rho = 0.40$	0.039	0.003	0.001	0.000
	$\rho = 0.80$	0.029	0.007	0.005	0.004

Table 6: Size and Power of the tests to detect for additive outliers: MA (1) errors

		$\delta_1 = 0, \delta_2 = 0,$ $\delta_3 = 0, \delta_4 = 0$			$\delta_1 = 5, \delta_2 = 3,$ $\delta_3 = 2, \delta_4 = 2$			$\delta_1 = 10, \delta_2 = 5,$ $\delta_3 = 5, \delta_4 = 5$		
	Probability to find	τ	τ_c	τ_d	τ	τ_c	τ_d	τ	τ_c	τ_d
$\theta = -0.80$	first outlier	0.166	0.242	0.056	0.834	0.865	0.746	0.998	0.999	1.000
	second outlier	0.026	0.001	0.002	0.360	0.121	0.179	0.957	0.828	0.921
	third outlier	0.004	0.000	0.000	0.094	0.001	0.019	0.865	0.359	0.779
	fourth outlier	0.001	0.000	0.000	0.022	0.000	0.002	0.661	0.147	0.518
$\theta = -0.40$	first outlier	0.063	0.098	0.054	0.286	0.342	0.941	0.793	0.823	1.000
	second outlier	0.021	0.003	0.002	0.058	0.009	0.396	0.401	0.199	0.997
	third outlier	0.011	0.000	0.000	0.013	0.000	0.076	0.194	0.021	0.985
	fourth outlier	0.007	0.000	0.000	0.007	0.000	0.007	0.081	0.004	0.912
$\theta = 0.00$	first outlier	0.040	0.065	0.047	0.101	0.135	0.996	0.464	0.516	1.000
	second outlier	0.023	0.004	0.002	0.024	0.004	0.674	0.122	0.035	1.000
	third outlier	0.015	0.000	0.000	0.014	0.000	0.228	0.038	0.001	1.000
	fourth outlier	0.010	0.000	0.000	0.010	0.000	0.040	0.013	0.000	0.998
$\theta = 0.40$	first outlier	0.036	0.054	0.038	0.055	0.079	1.000	0.231	0.282	1.000
	second outlier	0.024	0.004	0.004	0.022	0.003	0.821	0.044	0.007	1.000
	third outlier	0.018	0.000	0.001	0.015	0.000	0.380	0.016	0.001	1.000
	fourth outlier	0.012	0.000	0.000	0.010	0.000	0.106	0.010	0.000	1.000
$\theta = 0.80$	first outlier	0.036	0.053	0.021	0.042	0.063	0.994	0.130	0.164	1.000
	second outlier	0.026	0.004	0.005	0.023	0.003	0.749	0.026	0.004	1.000
	third outlier	0.019	0.000	0.001	0.017	0.000	0.297	0.015	0.000	1.000
	fourth outlier	0.014	0.000	0.001	0.012	0.000	0.087	0.011	0.000	1.000

Note: The Data Generating Process is: $y_t = \sum_{j=1}^{4} \delta_j D(T_{ao,j})_t + u_t$ with $u_t = u_{t-1} + v_t$ and $v_t = e_t + \theta e_{t-1}$ where $e_t \sim i.i.d.\ N(0,1)$. 10,000 replications are used.

Table 7: Size and Power of the tests to detect for additive outliers; AR (1)errors

		$\delta_1 = 0, \delta_2 = 0,$ $\delta_3 = 0, \delta_4 = 0$			$\delta_1 = 5, \delta_2 = 3,$ $\delta_3 = 2, \delta_4 = 2$			$\delta_1 = 10, \delta_2 = 5,$ $\delta_3 = 5, \delta_4 = 5$		
	Probability to find	τ	τ_c	τ_d	τ	τ_c	τ_d	τ	τ_c	τ_d
$\rho = -0.80$	first outlier	0.069	0.106	0.029	0.301	0.360	0.375	0.824	0.849	0.965
	second outlier	0.022	0.002	0.003	0.059	0.010	0.044	0.426	0.210	0.565
	third outlier	0.009	0.000	0.001	0.014	0.000	0.003	0.202	0.018	0.279
	fourth outlier	0.006	0.000	0.000	0.005	0.000	0.000	0.077	0.003	0.098
$\rho = -0.40$	first outlier	0.052	0.083	0.055	0.206	0.254	0.921	0.701	0.738	1.000
	second outlier	0.020	0.003	0.002	0.038	0.006	0.361	0.288	0.126	0.993
	third outlier	0.013	0.000	0.000	0.013	0.000	0.067	0.119	0.009	0.973
	fourth outlier	0.009	0.000	0.000	0.008	0.000	0.006	0.045	0.001	0.880
$\rho = 0.40$	first outlier	0.033	0.050	0.042	0.042	0.067	1.000	0.159	0.200	1.000
	second outlier	0.024	0.004	0.003	0.021	0.003	0.856	0.030	0.004	1.000
	third outlier	0.019	0.001	0.001	0.016	0.000	0.429	0.015	0.000	1.000
	fourth outlier	0.014	0.000	0.000	0.012	0.000	0.131	0.011	0.000	1.000
$\rho = 0.80$	first outlier	0.025	0.033	0.030	0.025	0.033	1.000	0.027	0.038	1.000
	second outlier	0.022	0.004	0.007	0.022	0.003	0.935	0.021	0.003	1.000
	third outlier	0.019	0.001	0.004	0.019	0.001	0.608	0.018	0.001	1.000
	fourth outlier	0.017	0.000	0.004	0.016	0.000	0.308	0.015	0.000	1.000

Note: The Data Generating Process is: $y_t = \sum_{j=1}^{4} \delta_j D(T_{ao,j})_t + u_t$ with $u_t = u_{t-1} + v_t$ and $v_t = \rho v_{t-1} + e_t$ where $e_t \sim i.i.d.\ N(0,1)$. 10,000 replications are used.

Table 8: Size of the tests τ_c and τ_d with non-unit root processes

		τ_c		τ_d	
		$T = 100$	$T = 200$	$T = 100$	$T = 200$
Near-Integrated Case	$c = -1.25$	0.062	0.065	0.053	0.051
	$c = -2.5$	0.065	0.071	0.053	0.051
	$c = -5.0$	0.083	0.091	0.054	0.051
	$c = -10.0$	0.126	0.145	0.054	0.051
	$c = -20.0$	0.206	0.248	0.055	0.050
	$c = -40.0$	0.281	0.362	0.056	0.051
i.i.d. Case		0.329	0.501	0.056	0.051

Note: For the Near-Integrated Case, the Data Generating Process is $y_t = (1 + c/T)y_{t-1} + e_t$ where $e_t \sim i.i.d.\ N(0, 1)$. For the *i.i.d.* Case, it is $y_t = e_t$. 10,000 replications are used.

Table 9: Size and power of τ_l and τ_d for stationary processes; MA (1) errors

		$\delta_1 = 0, \delta_2 = 0, \delta_3 = 0, \delta_4 = 0$		$\delta_1 = 5, \delta_2 = 3, \delta_3 = 2, \delta_4 = 2$		$\delta_1 = 10, \delta_2 = 5, \delta_3 = 5, \delta_4 = 5$	
	Probability to find	τ_l	τ_d	τ_l	τ_d	τ_l	τ_d
$\theta = -0.80$	first outlier	0.047	0.055	0.571	0.168	0.999	0.832
	second outlier	0.003	0.003	0.089	0.014	0.821	0.239
	third outlier	0.000	0.000	0.006	0.001	0.553	0.051
	fourth outlier	0.000	0.000	0.001	0.000	0.257	0.001
$\theta = -0.40$	first outlier	0.047	0.056	0.798	0.318	1.000	0.974
	second outlier	0.003	0.003	0.206	0.035	0.964	0.505
	third outlier	0.000	0.000	0.022	0.002	0.862	0.198
	fourth outlier	0.000	0.000	0.002	0.000	0.605	0.048
$\theta = 0.00$	first outlier	0.053	0.056	0.876	0.614	1.000	0.999
	second outlier	0.002	0.003	0.276	0.113	0.989	0.827
	third outlier	0.000	0.000	0.036	0.011	0.946	0.596
	fourth outlier	0.000	0.000	0.003	0.000	0.765	0.313
$\theta = 0.40$	first outlier	0.052	0.053	0.807	0.878	1.000	1.000
	second outlier	0.002	0.002	0.211	0.283	0.967	0.977
	third outlier	0.000	0.000	0.023	0.039	0.868	0.928
	fourth outlier	0.000	0.000	0.003	0.003	0.616	0.772
$\theta = 0.80$	first outlier	0.044	0.037	0.595	0.907	0.999	1.000
	second outlier	0.004	0.004	0.099	0.323	0.839	0.984
	third outlier	0.000	0.000	0.008	0.048	0.570	0.948
	fourth outlier	0.000	0.000	0.000	0.001	0.260	0.828

Note: The Data Generating Process is: $y_t = \sum_{j=1}^{4} \delta_j D(T_{ao,j})_t + u_t$ with $u_t = e_t + \theta e_{t-1}$ where $e_t \sim i.i.d.\ N(0, 1)$. 10,000 replications are used.

Table 10: Size and power of τ_l and τ_d for stationary processes; AR (1)errors

		$\delta_1=0,\delta_2=0,$ $\delta_3=0,\delta_4=0$		$\delta_1=0,\delta_2=0,$ $\delta_3=0,\delta_4=0$		$\delta_1=0,\delta_2=0,$ $\delta_3=0,\delta_4=0$	
	Probability to find	τ_l	τ_d	τ_l	τ_d	τ_l	τ_d
$\rho=-0.80$	first outlier	0.021	0.025	0.315	0.062	0.959	0.397
	second outlier	0.003	0.002	0.027	0.003	0.511	0.048
	third outlier	0.001	0.000	0.002	0.000	0.197	0.005
	fourth outlier	0.000	0.000	0.000	0.000	0.058	0.000
$\rho=-0.40$	first outlier	0.046	0.052	0.789	0.287	1.000	0.952
	second outlier	0.003	0.003	0.189	0.027	0.957	0.432
	third outlier	0.000	0.000	0.021	0.002	0.840	0.141
	fourth outlier	0.000	0.000	0.002	0.000	0.567	0.030
$\rho=0.40$	first outlier	0.051	0.056	0.804	0.885	1.000	1.000
	second outlier	0.003	0.003	0.200	0.283	0.963	0.979
	third outlier	0.000	0.000	0.022	0.043	0.855	0.938
	fourth outlier	0.000	0.000	0.002	0.004	0.591	0.788
$\rho=0.80$	first outlier	0.022	0.055	0.366	0.985	0.975	1.000
	second outlier	0.004	0.002	0.034	0.533	0.585	0.999
	third outlier	0.001	0.000	0.003	0.129	0.249	0.998
	fourth outlier	0.000	0.000	0.000	0.020	0.070	0.986

Note: The Data Generating Process is: $y_t = \sum_{j=1}^{4} \delta_j D(T_{ao,j})_t + u_t$ with $u_t = \rho u_{t-1} + e_t$ where $e_t \sim i.i.d.$ $N(0,1)$. 10,000 replications are used.

Table 11: Size of the ADF test; MA (1) errors
(choosing the lag length with the sequential t-sig method)

		$\delta_1=0,\delta_2=0,$ $\delta_3=0,\delta_4=0$			$\delta_1=5,\delta_2=3,$ $\delta_3=2,\delta_4=2$			$\delta_1=10,\delta_2=5,$ $\delta_3=5,\delta_4=5$		
		τ	τ_c	τ_d	τ	τ_c	τ_d	τ	τ_c	τ_d
$\theta=-0.80$	without	0.309	0.274	0.368	0.033	0.024	0.124	0.000	0.000	0.000
	with	0.089	0.122	0.021	0.399	0.402	0.309	0.380	0.460	0.392
	total	0.398	0.396	0.389	0.431	0.426	0.433	0.380	0.460	0.392
$\theta=-0.40$	without	0.069	0.062	0.081	0.040	0.031	0.007	0.003	0.001	0.000
	with	0.019	0.025	0.005	0.070	0.072	0.081	0.123	0.132	0.074
	total	0.088	0.087	0.086	0.110	0.103	0.088	0.125	0.133	0.074
$\theta=0.00$	without	0.038	0.034	0.049	0.048	0.041	0.001	0.011	0.008	0.000
	with	0.012	0.016	0.002	0.030	0.035	0.051	0.082	0.079	0.041
	total	0.050	0.049	0.051	0.078	0.076	0.052	0.094	0.087	0.041
$\theta=0.40$	without	0.048	0.044	0.057	0.035	0.029	0.000	0.024	0.017	0.000
	with	0.010	0.012	0.001	0.016	0.020	0.045	0.055	0.057	0.043
	total	0.057	0.056	0.058	0.051	0.049	0.045	0.079	0.074	0.043
$\theta=0.80$	without	0.048	0.045	0.056	0.036	0.031	0.000	0.044	0.037	0.000
	with	0.008	0.008	0.001	0.012	0.015	0.050	0.032	0.033	0.043
	total	0.056	0.053	0.057	0.047	0.046	0.050	0.076	0.070	0.043

Table 12: Size of the ADF test; AR (1) errors
(choosing the lag length with the sequential t-sig method)

		$\delta_1 = 0, \delta_2 = 0,$ $\delta_3 = 0, \delta_4 = 0$			$\delta_1 = 5, \delta_2 = 3,$ $\delta_3 = 2, \delta_4 = 2$			$\delta_1 = 10, \delta_2 = 5,$ $\delta_3 = 5, \delta_4 = 5$		
		τ	τ_c	τ_d	τ	τ_c	τ_d	τ	τ_c	τ_d
$\rho = -0.80$	without	0.040	0.034	0.049	0.034	0.028	0.061	0.001	0.001	0.008
	with	0.015	0.016	0.001	0.056	0.053	0.024	0.110	0.118	0.089
	total	0.055	0.051	0.050	0.090	0.081	0.085	0.111	0.119	0.097
$\rho = -0.40$	without	0.040	0.036	0.050	0.033	0.027	0.006	0.003	0.002	0.000
	with	0.013	0.017	0.003	0.042	0.043	0.047	0.099	0.097	0.045
	total	0.053	0.053	0.053	0.075	0.070	0.053	0.103	0.099	0.045
$\rho = 0.40$	without	0.041	0.038	0.050	0.031	0.026	0.000	0.020	0.015	0.000
	with	0.009	0.010	0.002	0.012	0.016	0.041	0.032	0.031	0.041
	total	0.050	0.048	0.052	0.043	0.042	0.041	0.052	0.046	0.041
$\rho = 0.80$	without	0.051	0.049	0.056	0.039	0.037	0.000	0.030	0.028	0.000
	with	0.004	0.006	0.001	0.004	0.006	0.045	0.004	0.007	0.043
	total	0.055	0.055	0.057	0.043	0.043	0.045	0.034	0.035	0.043

Table 13: Size of the ADF test; AR (1) errors
(with the lag length fixed at one)

		$\delta_1 = 0, \delta_2 = 0,$ $\delta_3 = 0, \delta_4 = 0$			$\delta_1 = 5, \delta_2 = 3,$ $\delta_3 = 2, \delta_4 = 2$			$\delta_1 = 10, \delta_2 = 5,$ $\delta_3 = 5, \delta_4 = 5$		
		τ	τ_c	τ_d	τ	τ_c	τ_d	τ	τ_c	τ_d
$\rho = -0.80$	without	0.037	0.031	0.046	0.056	0.046	0.091	0.013	0.009	0.019
	with	0.018	0.022	0.001	0.077	0.078	0.033	0.192	0.232	0.163
	total	0.055	0.053	0.047	0.133	0.124	0.124	0.205	0.241	0.182
$\rho = -0.40$	without	0.034	0.029	0.045	0.051	0.041	0.010	0.017	0.011	0.000
	with	0.016	0.021	0.002	0.058	0.058	0.057	0.159	0.177	0.050
	total	0.050	0.050	0.047	0.109	0.099	0.067	0.176	0.188	0.050
$\rho = 0.00$	without	0.034	0.030	0.046	0.035	0.030	0.000	0.021	0.015	0.000
	with	0.013	0.018	0.002	0.029	0.031	0.048	0.094	0.089	0.046
	total	0.047	0.048	0.048	0.064	0.061	0.048	0.115	0.104	0.046
$\rho = 0.40$	without	0.038	0.034	0.048	0.019	0.014	0.000	0.017	0.012	0.000
	with	0.009	0.013	0.002	0.011	0.015	0.041	0.023	0.025	0.045
	total	0.047	0.047	0.050	0.030	0.029	0.041	0.040	0.037	0.045
$\rho = 0.80$	without	0.047	0.045	0.052	0.023	0.021	0.000	0.028	0.025	0.000
	with	0.005	0.008	0.001	0.004	0.008	0.036	0.004	0.008	0.047
	total	0.052	0.053	0.053	0.027	0.029	0.036	0.032	0.033	0.047

Table 14: Empirical results; Logarithm of the US/Finland real exchange rate

Significance level	Test	CPI-based series 1900-1988	GDP-based series 1900-1987
5.0%	τ_c	1918	no outliers
	τ_d	1917,1918,1919,1932,1948	1918, 1949
10.0%	τ_c	1918,1919	no outliers
	τ_d	1917,1918,1919,1932,1948	1917,1918,1919,1921,1932,1947,1948,1957

Figure 1. Logarithm of the US/Finland Real Exchange Rate based on the Consumer Price Index (CPI); Annual from 1900 to 1988

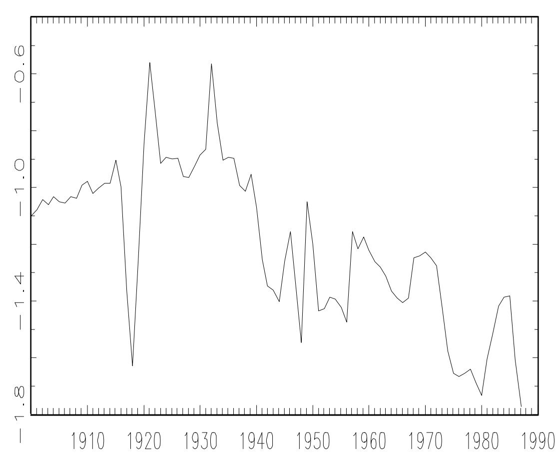

Figure 2. Logarithm of the US/Finland Real Exchange Rate based on the GDP De‡ator; Annual from 1900 to 1987