Changes in Persistence in Outlier Contaminated Time Series

Tristan Hirsch and Saskia Rinke[1]

Leibniz University Hannover

Abstract
Outlying observations in time series influence parameter estimation and testing procedures, leading to biased estimates and spurious test decisions. Further inference based on these results will be misleading. In this paper the effects of outliers on the performance of ratio-based tests for a change in persistence are investigated. We consider two types of outliers, additive outliers and innovative outliers. Our simulation results show that the effect of outliers crucially depends on the outlier type and on the degree of persistence of the underlying process. Additive outliers deteriorate the performance of the tests for high degrees of persistence. In contrast, innovative outliers do not negatively influence the performance of the tests. Since additive outliers lead to severe size distortions when the null hypothesis under consideration is described by a nonstationary process, we apply an outlier detection method designed for unit-root testing. The adjustment of the series results in size improvements and power gains. In an empirical example we apply the tests and the outlier detection method to the G7 inflation rates.

JEL-Numbers: C15, C22

Keywords: Additive Outliers · Innovative Outliers · Change in Persistence · Outlier Detection · Monte Carlo

[1]Corresponding Author. Leibniz University Hannover, School of Economics and Management, Institute of Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-Mail: `rinke@statistik.uni-hannover.de`. Phone: +49-511-762-3082. Fax: +49-511-762-3923.

1 Introduction

Since the introduction of additive outliers (AOs) and innovative outliers (IOs) by Fox (1972), the effect of outliers on statistical inference in time series has been investigated. Martin and Yohai (1986) consider the effect of outliers on parameter estimation. They show that isolated outliers induce a downward bias of the AR coefficients, whereas patches of outliers induce an upward bias. Franses and Haldrup (1994) assess the effect of AOs on the Dickey and Fuller (1979) unit-root test and find that the null hypothesis of a random walk is rejected too often (cf. also Shin et al., 1996). Besides, they also consider the Johansen (1991) trace test for cointegration and find cointegration too often. Hence, they conclude that AOs yield spurious stationarity as well as spurious cointegration and expect similar results in case of a temporary change. Also the performance of linearity tests is deteriorated in the presence of outliers and nonlinear models are preferred to linear models. According to van Dijk et al. (2002) this is due to the fact that nonlinear models can generate data resembling an outlier contaminated linear process. So, van Dijk et al. (1999) find that the test for smooth transition nonlinearity of Luukkonen et al. (1988) becomes oversized in the presence of AOs. In extreme scenarios the size distortion improves but power losses occur. In contrast, IOs do not seriously deteriorate the performance of the test. Therefore, they conclude that the influence of AOs is much more severe than the effects of IOs. Ahmad and Donayre (2016) find evidence for size distortions but power improvements due to outliers for the test against threshold autoregressive nonlinearity of Hansen (1996, 1997).

The effect of outliers on tests for a change in persistence has not been assessed yet. Therefore, in this paper we investigate the performance of the ratio-based tests of Kim (2000); Kim et al. (2002) and of Leybourne et al. (2007) in outlier contaminated processes. Both tests are based on a ratio of the subsample cumulative sum of squared residuals. Outliers influence the test statistic via the residuals and thus can lead to spurious test decisions.

In our simulation studies we vary the outlier magnitude, the sample size, and the change magnitude to assess their individual effects. Furthermore, we apply the outlier detection method of Shin et al. (1996) which is designed for unit-root testing and compare the performance of the tests in the contaminated and in the adjusted series.

The rest of the paper is organized as follows. In Section 2 the model framework and the different outlier types are introduced. In Section 3 the tests for a change in persistence are explained. Section 4 introduces the outlier detection and removal methods. In Section 5 the simulation set-up and the simulation results are presented. Section 6 contains a real data example of the G7 inflation rates. Finally, Section 7 concludes.

2 Modeling Outliers and Changes in Persistence

Outliers can only be defined in the context of a certain model under consideration (cf. Davies and Gather, 1993; van Dijk et al., 1999). In our analysis we will focus on autoregressive processes of order 1 with and without a change in persistence,

$$\Phi(L)\,x_t = \varepsilon_t, \qquad\qquad t = 1,\ldots,T, \tag{2.1}$$

where T is the sample size, $\Phi(L) = 1 - \phi_1 L\mathbf{1}\{t \le \lfloor\tau\cdot T\rfloor\} - \phi_2 L\mathbf{1}\{t > \lfloor\tau\cdot T\rfloor\}$, L is the lag operator, $\mathbf{1}\{\cdot\}$ is the indicator function, $\lfloor\tau\cdot T\rfloor$ is the change point, and $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$. There is no change in persistence if $\phi_1 = \phi_2$, $\tau = 0$, or $\tau = 1$. A common way to model outliers in the context of linear time series is the general replacement model of Martin and Yohai (1986),

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

The observable contaminated process y_t consists of the unobservable core process x_t and the contaminating process ζ_t. The random variable δ_t takes the values -1 and 1, each with the probability $\pi/2$, and 0 otherwise, where the probability π is the outlier probability. Allowing δ_t to take positive and negative values, enables us to model symmetric contaminations. The core process is the AR(1) model of Eq. (2.1).

Depending on the specification of the contaminating process ζ_t, different types of outliers are generated, i.e. AOs, IOs, level shifts, and temporary changes (cf. Galeano and Peña, 2013). In the context of time series mostly AOs and IOs are considered (cf. Fox, 1972; van Dijk et al., 1999). For AOs the contaminating process ζ_t and the respective contaminated process y_t are given by

$$\zeta_t = x_t + \zeta,$$
$$\text{and} \quad y_t = x_t + \zeta\delta_t,$$

where ζ is the constant outlier magnitude depending on the standard deviation of the core process, σ_x. An IO contamination ζ_t and its observable process y_t can be modeled as

$$\zeta_t = x_t + \zeta/\Phi(L)$$
$$\text{and} \quad y_t = x_t + \big(\zeta/\Phi(L)\big)\,\delta_t.$$

AOs only have a one-time effect on the series since they do not affect the core process x_t. In contrast IOs have a one-time effect on the errors but influence several observations through the dynamics of the core process. Therefore, IOs have different effects in stationary and in nonstationary core processes. In contrast to IOs in stationary processes, the effect of an IO in a unit-root process is permanent and similar to a level shift.

3 Tests for a Change in Persistence

Several procedures exist to test for a change in persistence. They include the ratio-based tests of Kim (2000), Kim et al. (2002), Busetti and Taylor (2004), and Leybourne et al. (2007) among others, the sub-sample augmented Dickey-Fuller-type test of Leybourne et al. (2003), and the variance ratio test of Leybourne et al. (2004). All tests assume a constant persistence under the null hypothesis, either $I(0)$ like in Kim (2000) or $I(1)$ like in Leybourne et al. (2007). The alternative is a change from $I(0)$ to $I(1)$ ($I(0) \to I(1)$) or a change from $I(1)$ to $I(0)$ ($I(1) \to I(0)$). We will focus on the test of Kim (2000); Kim et al. (2002) (the Kim test) since it is frequently applied and on the test of Leybourne et al. (2007) (the Leybourne test) due to its good size and power properties. The idea of the tests is to divide the time series into two subsamples and take the ratio of the subsample cumulative sum (CUSUM) of squared residuals. For both tests simulated critical values are tabulated for the relevant sample sizes and significance levels of the simulation study in Section 5.

3.1 The Kim Test

Kim (2000) and Kim et al. (2002) test the null hypothesis of constant $I(0)$ against a change in persistence $I(0) \to I(1)$ with the test statistic

$$K_{\lfloor \tau T \rfloor} = \frac{(T - \lfloor \tau T \rfloor)^{-2} \sum_{t=\lfloor \tau T \rfloor+1}^{T} \left(\sum_{i=\lfloor \tau T \rfloor+1}^{t} \tilde{v}_{i,\tau} \right)^2}{\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\sum_{i=1}^{t} \hat{v}_{i,\tau} \right)^2},$$

where $\hat{v}_{t,\tau}$ are the residuals from the OLS regression of y_t on a constant term for observations up to $\lfloor \tau T \rfloor$ to obtain invariance to a constant. Similarly, $\tilde{v}_{t,\tau}$ are the OLS residuals from the regression of y_t on a constant term for $t = \lfloor \tau T \rfloor + 1, \ldots, T$. Since the true change point τ^* is unknown, Kim (2000), Kim et al. (2002), and Busetti and Taylor (2004) use the sequence of statistics $\{K_{\lfloor \tau T \rfloor}\}$ for $\tau \in \Lambda$, where the change fraction τ^* is assumed to lie in $\Lambda = [\tau_l, \tau_u]$, an interval in $(0,1)$ which is symmetric around 0.5, typically $[0.2, 0.8]$. Following Leybourne et al. (2007) we will only consider the maximum test. Then, the test statistic and the estimated change fraction are given by

$$MX = \max_{\tau \in \Lambda} K_{\lfloor \tau T \rfloor},$$
$$\hat{\tau} = \arg\sup_{\tau \in \Lambda} \Xi(\tau),$$

with $\Xi(\tau) = \left((T - \lfloor \tau T \rfloor)^{-2} \sum_{t=\lfloor \tau T \rfloor + 1}^{T} \tilde{v}_{i,\tau}^2 \right) \left(\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} \hat{v}_{i,\tau}^2 \right)^{-1}$. The null hypothesis will be rejected if the value of the test statistic MX is smaller or larger than the lower or upper tail critical value, respectively.

In Table 3.1 simulated upper and lower tail critical values of the Kim test for different sample sizes are given. They are based on 100 000 replications.

	Quantile					
T	0.005	0.025	0.050	0.950	0.975	0.995
50	0.534	0.910	1.185	16.878	21.588	35.050
100	0.594	0.992	1.292	17.047	21.591	34.001
250	0.647	1.087	1.402	17.776	22.425	36.033
500	0.681	1.111	1.438	17.932	22.646	35.489
1000	0.679	1.140	1.475	18.202	23.084	36.036

Table 3.1: Simulated Critical Values of the Kim Test

3.2 The Leybourne Test

In contrast to the Kim test, Leybourne et al. (2007) test the null hypothesis of constant $I(1)$ against a change in persistence from $I(0) \rightarrow I(1)$ or $I(0) \rightarrow I(1)$ with the following two-tailed test statistic

$$R = \frac{K^f(\tau)}{K^r(\tau)} = \frac{\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} \hat{v}_{t,\tau}^2}{(T - \lfloor \tau T \rfloor)^{-2} \sum_{t=1}^{(T - \lfloor \tau T \rfloor)} \tilde{v}_{t,\tau}^2}, \tag{3.1}$$

where $K^f(\tau)$ is the forward test statistic with $\hat{v}_{t,\tau}$ as defined above and $K^r(\tau)$ is the test statistic for the reversed series. Note that a change $I(1) \rightarrow I(0)$ is equivalent to a change $I(0) \rightarrow I(1)$ in the reversed series, $\tilde{y}_t \equiv y_{T-t+1}$, occurring at time $T - \lfloor \tau^* T \rfloor$.

Leybourne et al. (2007) show that $K^f(\tau)$ converges in probability to zero for a change $I(0) \rightarrow I(1)$ for all $\tau \leq \tau^*$ and is of $O_p(1)$ if the persistence changes from $I(1) \rightarrow I(0)$ for all τ. $K^r(\tau)$ converges in probability to zero if $I(1) \rightarrow I(0)$ for all $\tau > \tau^*$ and is of $O_p(1)$

if $I(0) \to I(1)$ for all τ. So, if the true change point $\tau^* T$ is known, a test of the null hypothesis $I(1)$ against a change in persistence, either $I(0) \to I(1)$ or $I(1) \to I(0)$, can be based on Eq. (3.1), because a ratio of $K^f(\tau^*)$ and $K^r(\tau^*)$ collapses to zero for $I(0) \to I(1)$ and diverges to positive infinity for $I(1) \to I(0)$. Because the true change fraction τ^* is unknown, the test is based on the infima of $K^f(\tau)$ and $K^r(\tau)$ for $\tau \in \Lambda$. The null hypothesis of $I(1)$ throughout will be rejected if R exceeds or falls below the upper or the lower tail critical value, respectively. The estimated change fraction $\hat{\tau}$ is given by $\arg\inf_{\tau \in \Lambda} K^f(\tau)$ for a change $I(0) \to I(1)$ and by $\arg\inf_{\tau \in \Lambda} K^r(\tau)$ for a change $I(1) \to I(0)$. In Table 3.2 simulated upper and lower tail critical values of the Leybourne test for different sample sizes are given. They are based on 100 000 replications.

	Quantile					
T	0.005	0.025	0.050	0.950	0.975	0.995
50	0.131	0.213	0.276	3.600	4.686	7.616
100	0.117	0.194	0.256	3.950	5.149	8.572
250	0.104	0.180	0.239	4.177	5.502	9.531
500	0.100	0.177	0.234	4.278	5.684	10.017
1000	0.101	0.177	0.234	4.327	5.773	10.152

Table 3.2: Simulated Critical Values of the Leybourne Test

Leybourne et al. (2007) show that the test is conservative against a constant $I(0)$ process. Thus, in contrast to the Kim test the Leybourne test does not spuriously detect changes in persistence.

4 Outlier Detection and Removal Methods

There are several publications emphasizing the deteriorating effect of outliers on the performance of estimation and testing methods (cf. Franses and Haldrup (1994); van Dijk et al. (1999); Ahmad and Donayre (2016) among others). Two strands of procedures exist in order to handle outlier contaminated series. Either the outliers have to be detected and removed before parameters are estimated and tests are conducted, or the approaches have to be robust against outliers (cf. e.g. van Dijk et al., 1999). Several outlier detection methods have been proposed starting with Chang et al. (1988) and Tsay (1988). The approach of Tsay (1988) works under the initial assumption of an uncontaminated series and consists of specification and estimation in an outer loop and detection and removal of outliers in the inner loop (cf. Figure 4.1). In a first step the critical value C as well as the order of an ARMA model have to be selected and the corresponding parameters are estimated. The inner loop starts with the calculation of the residuals and the estimation of the error term variance $\hat{\sigma}_\varepsilon^2$. For each outlier type $j = AO, IO$ and each observation $t = 1, \ldots, T$ the test statistic $\lambda_{j,t} = \hat{\zeta}_{j,t}/\hat{\sigma}_j$, where $\hat{\zeta}_{j,t}$ is the estimated outlier effect and $\hat{\sigma}_j$ is the corresponding standard deviation depending on $\hat{\sigma}_\varepsilon$, is calculated to test the null hypothesis of no outlier of type j at observation t,

$$H_0 : \hat{\zeta}_{j,t} = 0 \qquad\qquad H_1 : \hat{\zeta}_{j,t} \neq 0.$$

Let t_j denote the observation with the highest probability of being an outlier of type j. In order to identify t_j, Tsay (1988) takes the maximum of the test statistics $\lambda_{j,t}$ over all t. The maximum of both $\lambda_{AO,t_{AO}}$ and $\lambda_{IO,t_{IO}}$ denotes the final test statistic λ to determine the outlier type and position. If λ exceeds the critical value C the outlier is removed depending on the type and the inner loop further iterates.

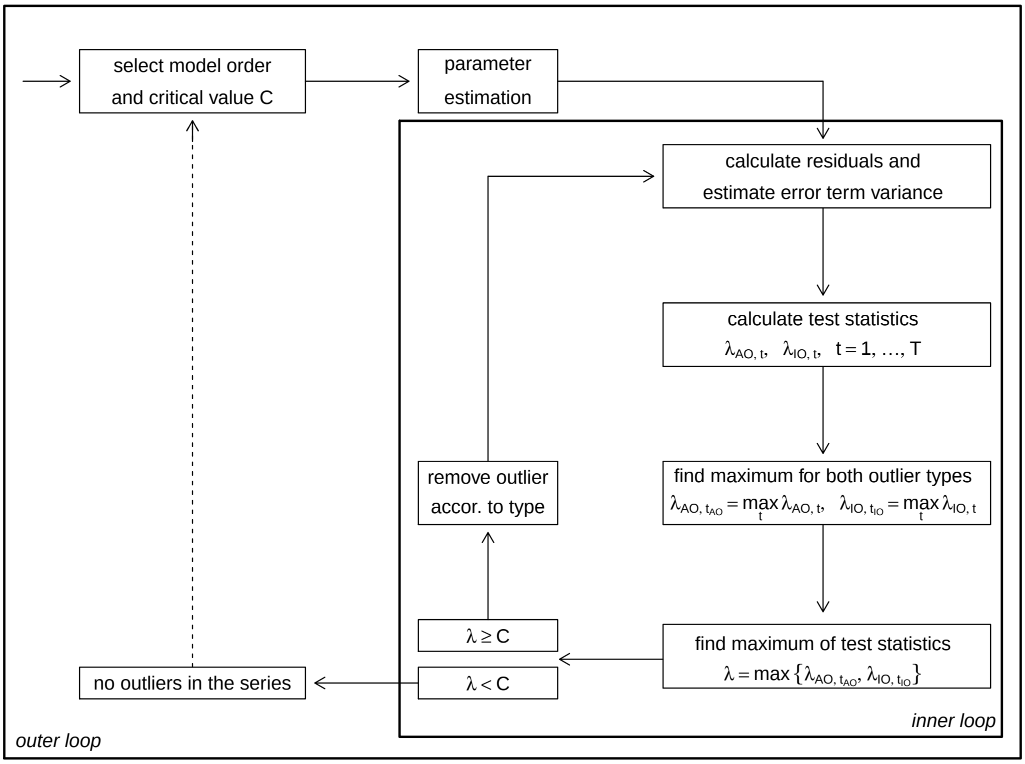


Figure 4.1: The Outlier Detection and Removal Method of Tsay (1988)

If the inner loop is completed after one single iteration, the algorithm stops and the series is uncontaminated. If however the inner loop stops after iterating several times to remove outliers, the outer loop starts again to check a refined model.

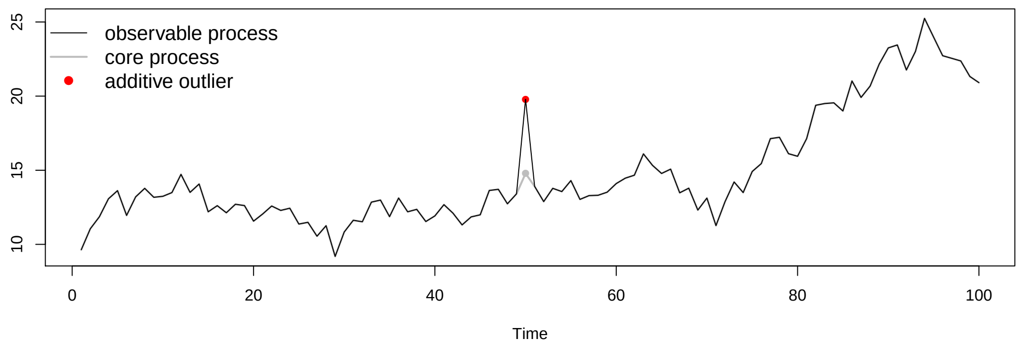
The described algorithm detects outliers sequentially, which is computationally easier and performs well if there exists only a single outlier in the series but can lead to biased estimates if there are multiple outliers (cf. Chen and Liu, 1993). Therefore, Chen and Liu (1993) propose a procedure consisting of three different stages.

In the first stage the algorithm of Tsay (1988) is applied to detect possible outliers. Given the information of the first stage about the estimated time points where outliers occur, the outlier effects can be estimated jointly and the significance of the outliers is assessed. Insignificant outliers are deleted one-by-one until all remaining outlier effects are significant. Finally the model parameters are estimated. Given this information, in the third stage the procedure starts again with the refined parameter estimates.

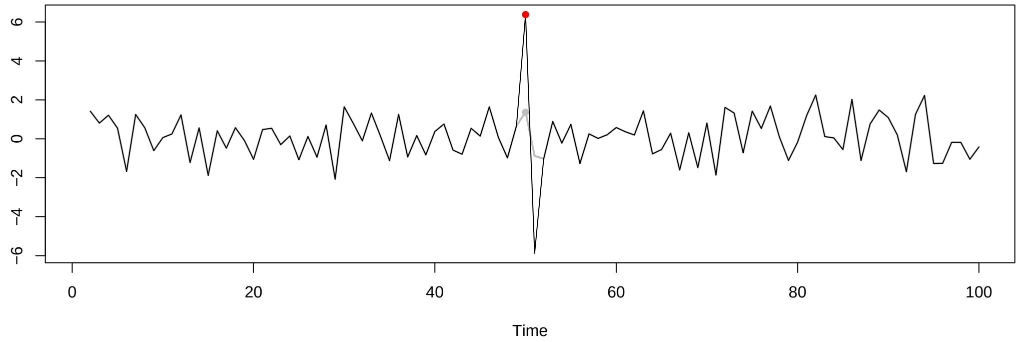
According to Galeano and Peña (2013) the procedure of Chen and Liu (1993) is the standard approach for outlier detection in linear time series. However, it has three major drawbacks, firstly, the type of outlier (IO or level shift) may not be correctly identified which affects the adjustment of the series, secondly, the algorithm depends on initial parameter estimates, may resulting in the break down of the procedure due to biased initial values, and finally, patches of outliers may not be identified due to the masking effect. Sánchez and Peña (2003) further modify the approach in order to solve these problems. For example, they calculate robust initial estimates by eliminating influential points (cf. also Peña, 1991) and use lower critical values C to be able to identify patches of outliers. Although further extensions lead to improved results, the computational burden increases enormously. Moreover, the main aim of the detection algorithms is to obtain unbiased parameter estimates for an ARMA model.

Since we are primarily interested in the demeaned series, we apply the algorithm of Shin et al. (1996) which focuses on outlier detection for unit-root testing and works under the assumption of the series being a random walk. This approach can be valuable in our analysis, since the test by Leybourne et al. (2007) is $I(1)$ under the null hypothesis. However, the test of Kim (2000); Kim et al. (2002) is $I(0)$ under the null hypothesis and Shin et al. (1996) admit that their outlier detection algorithm does not perform well if the process under consideration exhibits only a small degree of persistence. Nevertheless, our results in the simulation studies show that the performance of the Kim test is not deteriorated by outliers if the process only exhibits a low degree of persistence. Due to the assumption of a random walk, the procedure of Shin et al. (1996) does not need an initial model selection and parameter estimates, thus minimizing the computational effort.

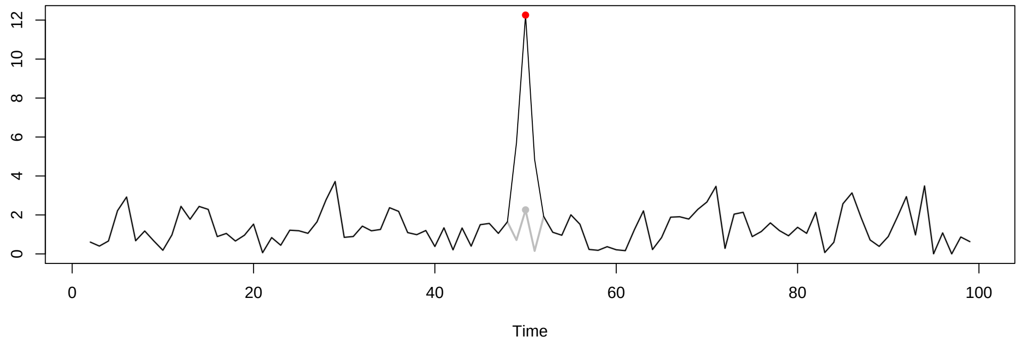
The idea of the Shin et al. (1996) algorithm is illustrated in Figure 4.2. An AO only affects one single observation but two consecutive residuals, i.e. the differences between two consecutive observations, $e_t = y_t - y_{t-1}$. Thus, a test can be based on the difference between the residuals. Since the difference may be negative, the absolute value is considered.



(a) Random Walk with an Additive Outlier at $t = 50$



(b) Residuals as the First Difference of the Random Walk



(c) Absolute value of the First Difference of Residuals

Figure 4.2: Idea of the Shin et al. (1996) Algorithm

Due to the fact that it is not known a priori when an AO occurs, the maximum of the absolute differences is determined. Let $t_{AO} = \arg\max_{2 \le t \le T-1} \left| e_{t+1} - e_t \right|$, then t_{AO} is the observation that is most likely to be contaminated by an AO. To test whether there occurs an AO at t_{AO}, $|e_{t_{AO}+1} - e_{t_{AO}}|$ is standardized by the estimated standard deviation of $e_{t_{AO}+1} - e_{t_{AO}}$. The general test statistic is given by

$$\lambda = \frac{1}{\sqrt{2}\hat{\sigma}} \left(\max_{2 \le t \le T-1} |e_{t+1} - e_t| \right),$$

where $\hat{\sigma}^2 = (T-3)^{-1} \left((\sum_{t=2}^{T} e_t^2) - e_{t_{AO}}^2 - e_{t_{AO}+1}^2 \right)$ is a robust estimator of σ_ε^2. If the test statistic equals or exceeds a critical value C, an AO is detected. We follow Shin et al. (1996) and use the critical value $C = 3$. A further discussion of the distribution of λ can be found in the appendix.

Shin et al. (1996) recommend to replace an AO contaminated observation with its lagged value to adjust the series. This procedure only takes into account the information up to t_{AO} and leads to constant parts in the time series, resulting in a larger residual $e_{t_{AO}+1}$. Therefore, we suggest to use the full sample information and to replace the outlying observation $y_{t_{AO}}$ by its best full sample prediction, i.e. the mean of the lagged value and the future value, $\hat{y}_{t_{AO}} = (y_{t_{AO}-1} + y_{t_{AO}+1})/2$. The procedure is repeated until no additional outliers are detected, i.e. $\lambda < C$.

The approach can be adjusted to detect IOs (cf. Shin et al., 1996). However, as we will show in the following section, this is not necessary, since IOs do not seriously affect the performance of the tests for a change in persistence.

5 Simulation Study

In our simulation study we consider the linear model given in Eq. (2.1) without contaminations ($\zeta = 0$) and with AOs as well as IOs of different outlier magnitudes ζ with an outlier probability of $\pi = 0.05$ (cf. Ahmad and Donayre, 2016). The errors form a Gaussian white noise process. In order to assess the performance of the tests, we apply them to the uncontaminated, contaminated, and adjusted series. To adjust the series we use the modified algorithm of Shin et al. (1996) with a critical value of $C = 3$. We vary the following parameters,

$$\begin{array}{lcl} \text{sample size} & T & = \{50, 100, 250, 500, 1000\}, \\ \text{persistence} & \phi_1, \phi_2 & = \{0.00, 0.25, 0.50, 0.75, 0.95, 1.00\}, \\ \text{outlier magnitude} & \zeta & = \{0\sigma_x, 1\sigma_x, 2\sigma_x, 3\sigma_x\}. \end{array}$$

For every series 200 additional observations are simulated as a burn-in period to avoid a starting value bias. All initial values are set to zero. The simulation results are based on 1000 replications. The following figures and tables report the simulation results for $\tau = 0.5$. In general we find that the power of the tests is higher if the change point occurs early in the series under the condition that the stationary part of the series is at least as large as the nonstationary part.

5.1 Performance in Uncontaminated Series

Table 5.1 tabulates the size properties of the Kim and the Leybourne test in uncontaminated series for different sample sizes T and different levels of significance α.

	significance level α		
T	1%	5%	10%
50	0.009	0.049	0.093
100	0.011	0.052	0.110
250	0.009	0.045	0.094
500	0.008	0.048	0.095
1000	0.010	0.050	0.093

(a) Kim Test $I(0)$

	significance level α		
T	1%	5%	10%
50	0.011	0.046	0.105
100	0.011	0.047	0.097
250	0.010	0.042	0.094
500	0.010	0.052	0.107
1000	0.009	0.046	0.102

(b) Leybourne Test $I(1)$

Table 5.1: Size Properties

The size of the Kim and of the Leybourne test coincides with the nominal size. Since the critical values depend on the number of observations, the tests perform well in terms of size for all sample sizes.
Table 5.2 tabulates the power results of the Kim test for $I(0) \rightarrow I(1)$ and of the Leybourne test for both $I(0) \rightarrow I(1)$ and $I(1) \rightarrow I(0)$.

	significance level α		
T	1%	5%	10%
50	0.779	0.868	0.907
100	0.947	0.978	0.982
250	0.997	0.998	0.999
500	1.000	1.000	1.000
1000	1.000	1.000	1.000

(a) Kim Test $I(0) \rightarrow I(1)$

	significance level α		
T	1%	5%	10%
50	0.017	0.081	0.156
100	0.084	0.262	0.400
250	0.408	0.690	0.803
500	0.798	0.935	0.977
1000	0.963	0.997	0.999

(b) Leybourne Test $I(1) \rightarrow I(0)$

	significance level α		
T	1%	5%	10%
50	0.087	0.246	0.393
100	0.238	0.504	0.666
250	0.612	0.858	0.934
500	0.882	0.979	0.995
1000	0.987	1.000	1.000

(c) Leybourne Test $I(0) \rightarrow I(1)$

Table 5.2: Power Properties

The power of both tests increases with the sample size. However, in small samples the power of the Kim test is already high and it converges to 1 with an increasing number of observations. In contrast, the power of the Leybourne test crucially depends on the sample size. In very small samples $T = 50$ the power is only slightly higher than its size. Also for $T = 100$ the power is relatively low. For sample sizes of $T \geq 250$ the power increases and the test decision is reliable. With an increasing number of observations the power of the test converges to 1.

All presented results are valid for $\phi_1, \phi_2 = \{0, 1\}$. In general, the size of the Kim test increases if the degree of persistence increases and the power decreases with a decreasing change magnitude $|\phi_1 - \phi_2|$ (cf. Fig. A.3 and A.4). For the Leybourne test the size decreases to zero if the process becomes stationary. The power decreases if $|\phi_1 - \phi_2|$ decreases (cf. Fig. A.5 and A.6).

5.2 Performance in Contaminated Series

Figure 5.1 illustrates the effects of AOs and IOs on the size of the Kim test for different sample sizes, outlier magnitudes ζ, and significance levels. The results show that there is no difference between the effects of AOs and IOs on the size of the Kim test. This is due to the fact that the degree of persistence of the core process is zero under the null hypothesis and an IO can only affect one observation exactly like an AO. The effect of outliers is mostly pronounced for large outlier magnitudes ζ and small to moderate sample sizes. The higher the persistence of the simulated processes, the higher are the size distortions in small samples (cf. Fig. A.3 and A.4). However, the size is not deteriorated seriously, but holds the nominal significance level.

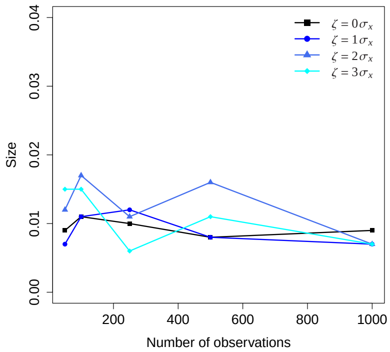
The power of the Kim test is not affected by AO contaminations if ζ is small. Only for large outlier magnitudes $\zeta = 3\sigma_x$ the power of the test decreases in small samples. The power of the test is not affected by IO contaminations (cf. Fig. 5.2).

Figure 5.3 presents the size of the Leybourne test in outlier contaminated series for different sample sizes, outlier magnitudes, and levels of significance. In the left panel the results for AOs can be found. The introduction of AOs decreases the size of the Leybourne test for all sample sizes and all significance levels. This implies that the test becomes undersized. The size distortion increases with the sample size and the outlier magnitude. For large sample sizes combined with large outlier magnitudes the size converges to zero. This is due to the fact that an AO contaminated unit-root process can be confused with a stationary process (cf. Franses and Haldrup, 1994). Since the size of the Leybourne test converges to zero for a constant $I(0)$ process, the size of the Leybourne test decreases to zero in AO contaminated series.

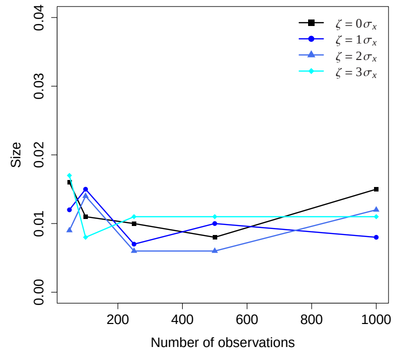
In the right panel of Figure 5.3 the size properties of the Leybourne test in IO contaminated time series are depicted. The size distortions are less severe compared to AO contaminations (cf. also van Dijk et al., 1999). Only in small samples and for large outlier magnitudes the size differs from the nominal significance level.

In terms of size the Leybourne test is more affected by outliers than the Kim test due to the higher degree of persistence under the null hypothesis. The effect of AOs is more serious than the effect of IOs.

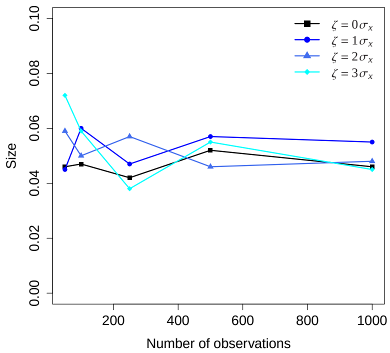
Figures 5.4 and 5.5 illustrate the power properties of the Leybourne test for $I(0) \to I(1)$ and $I(1) \to I(0)$, respectively. For both alternatives the results are qualitatively the same. For a change $I(0) \to I(1)$ the power is slightly higher across sample sizes, significance levels, and outlier magnitudes. This coincides with the findings in the uncontaminated series (cf. Tab. 5.2). In the left panels the effects of AOs on the power properties are depicted. The power decreases and approaches zero for increasing outlier magnitudes because the contaminated series can be confused with a stationary $I(0)$ process. In contrast, IOs do not decrease the power, but lead to power gains since the stationary and the nonstationary part of the series markedly differ (cf. Fig. 5.6a).



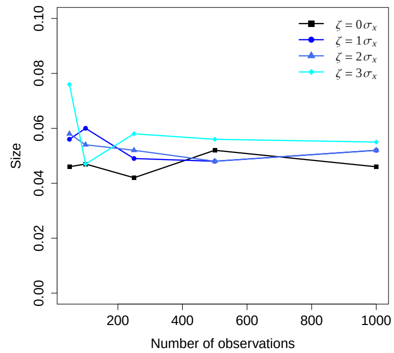
(a) AOs and $\alpha = 1\%$



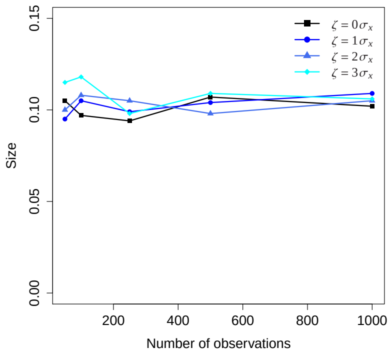
(b) IOs and $\alpha = 1\%$



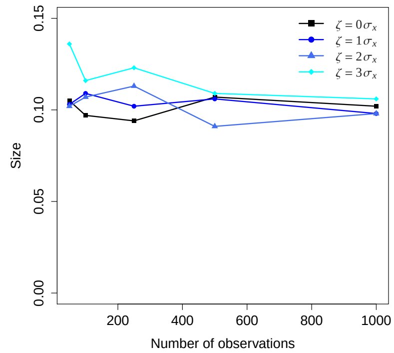
(c) AOs and $\alpha = 5\%$



(d) IOs and $\alpha = 5\%$

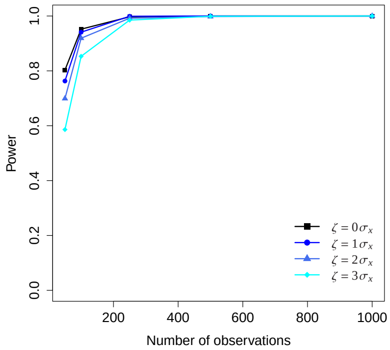


(e) AOs and $\alpha = 10\%$

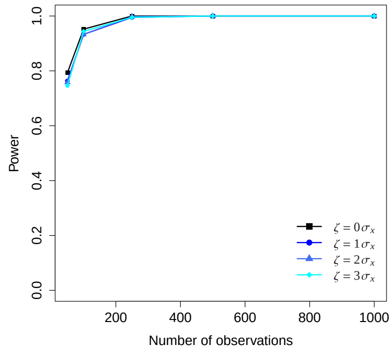


(f) IOs and $\alpha = 10\%$

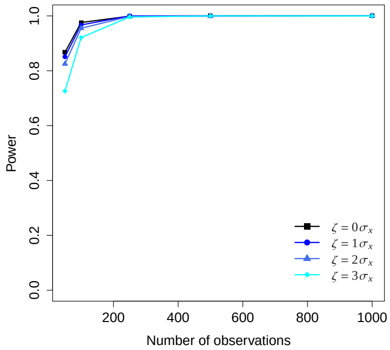
Figure 5.1: Size of the Kim Test ($I(0)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance



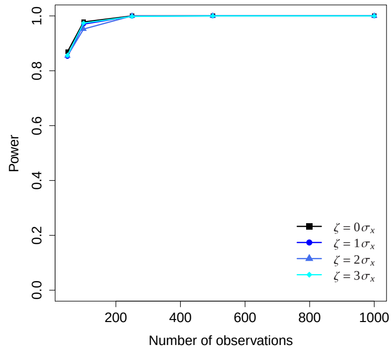
(a) AOs and $\alpha = 1\%$



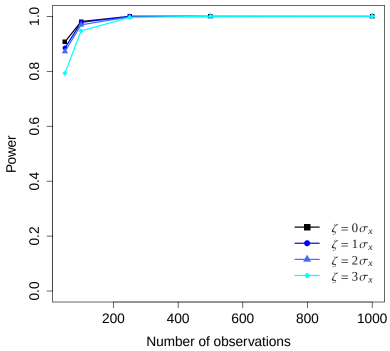
(b) IOs and $\alpha = 1\%$



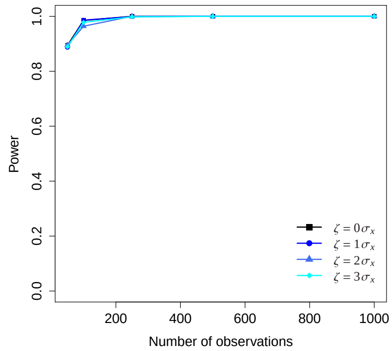
(c) AOs and $\alpha = 5\%$



(d) IOs and $\alpha = 5\%$

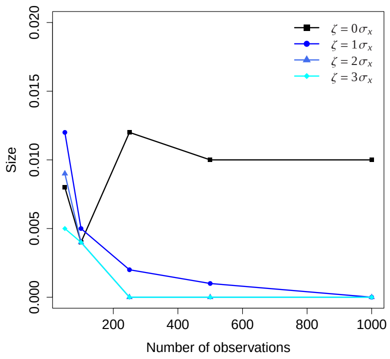


(e) AOs and $\alpha = 10\%$

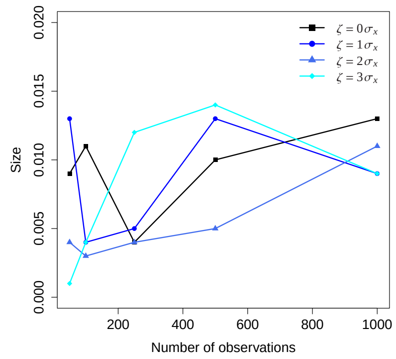


(f) IOs and $\alpha = 10\%$

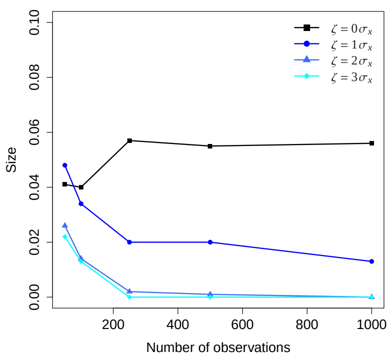
Figure 5.2: Power of the Kim Test ($I(0) \rightarrow I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance



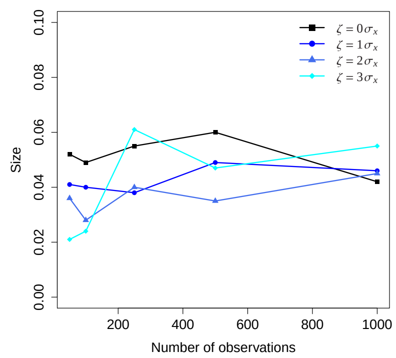
(a) AOs and $\alpha = 1\%$



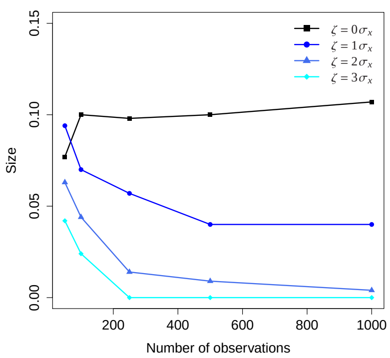
(b) IOs and $\alpha = 1\%$



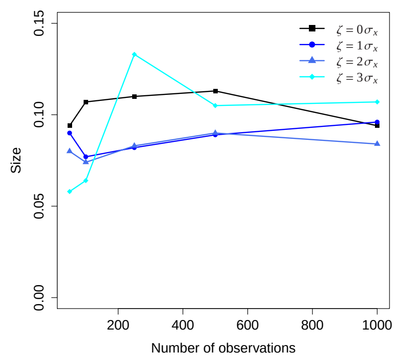
(c) AOs and $\alpha = 5\%$



(d) IOs and $\alpha = 5\%$

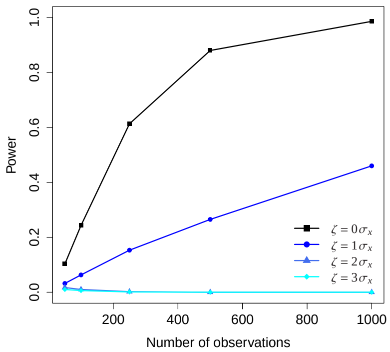


(e) AOs and $\alpha = 10\%$

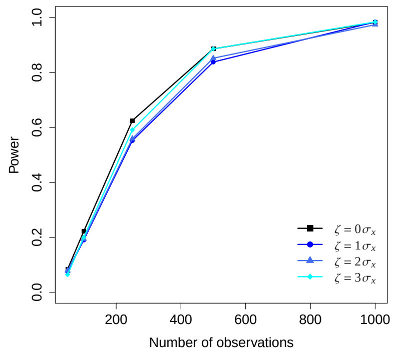


(f) IOs and $\alpha = 10\%$

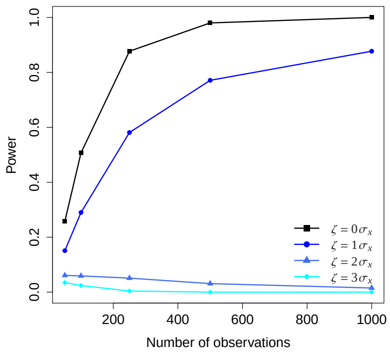
Figure 5.3: Size of the Leybourne Test ($I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance



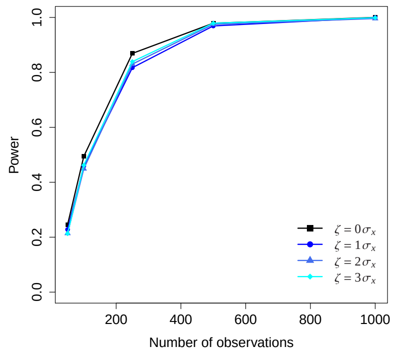
(a) AOs and $\alpha = 1\%$



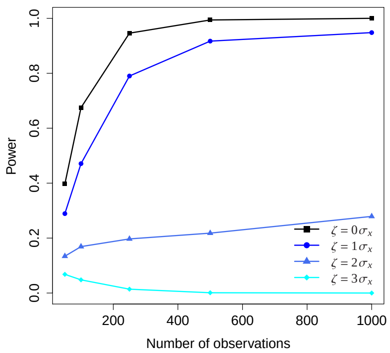
(b) IOs and $\alpha = 1\%$



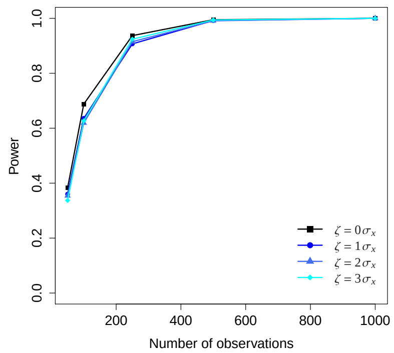
(c) AOs and $\alpha = 5\%$



(d) IOs and $\alpha = 5\%$

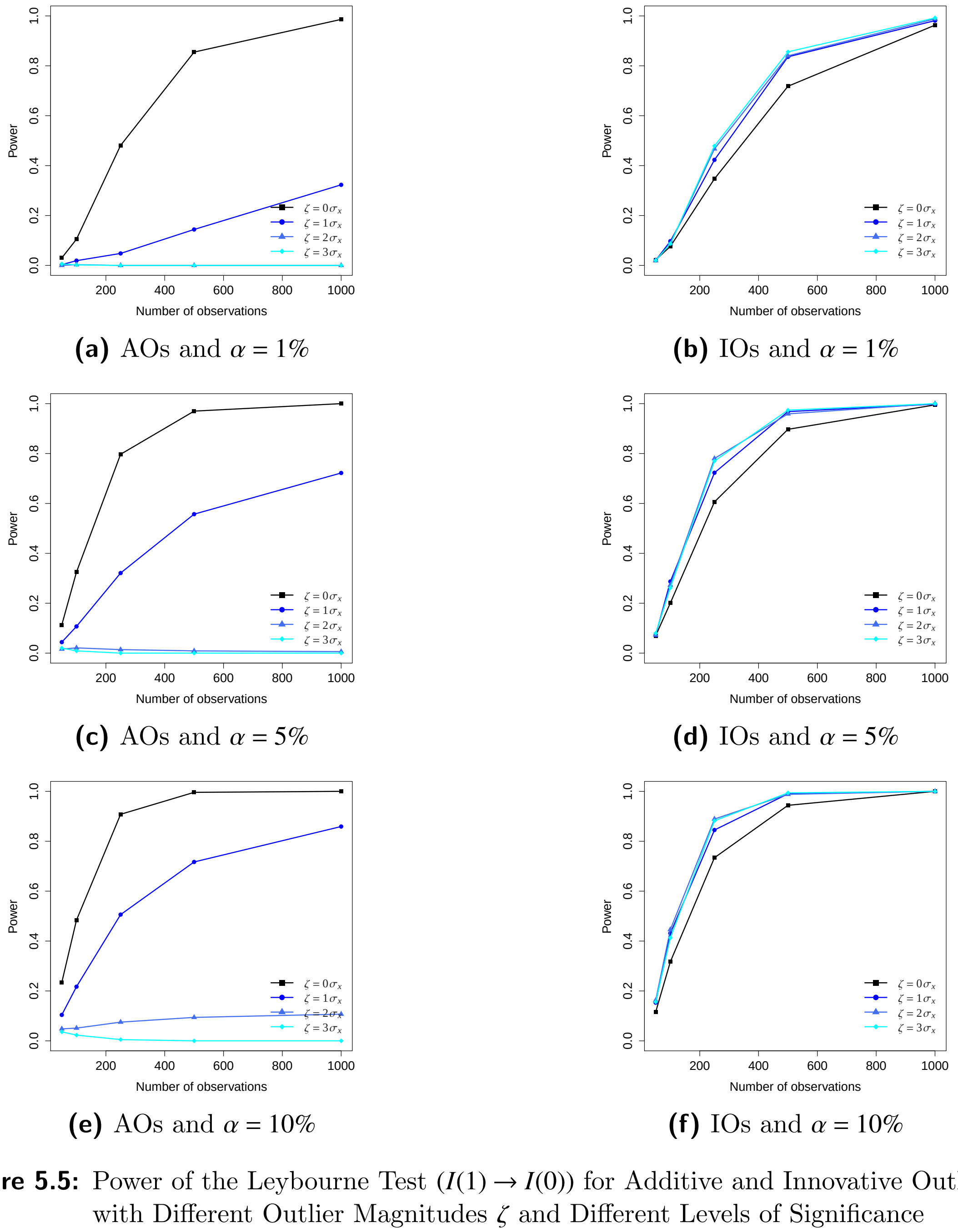


(e) AOs and $\alpha = 10\%$



(f) IOs and $\alpha = 10\%$

Figure 5.4: Power of the Leybourne Test ($I(0) \rightarrow I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance



(a) AOs and $\alpha = 1\%$

(b) IOs and $\alpha = 1\%$

(c) AOs and $\alpha = 5\%$

(d) IOs and $\alpha = 5\%$

(e) AOs and $\alpha = 10\%$

(f) IOs and $\alpha = 10\%$

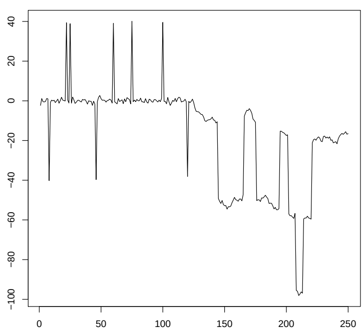
Figure 5.5: Power of the Leybourne Test ($I(1) \rightarrow I(0)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance

5.3 Performance in Adjusted Series

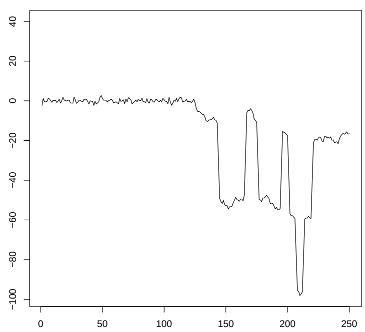
The results in Figures 5.1 and 5.2 show that the performance of the Kim test does not suffer from size distortions or power losses due to outliers for low degrees of persistence. Hence, it is not necessary to adjust the series before applying the test. Moreover, the modified algorithm of Shin et al. (1996) is developed for nonstationary time series and thus does not perform well in series with a low degree of persistence. Although the application of the Kim test to the adjusted series results in power gains, it also suffers from an increased size (cf. Fig. A.3 and A.4).

Figure 5.7 shows the size properties of the Leybourne test in the adjusted series. In all uncontaminated series the size is not affected by the adjustment procedure. Therefore, the algorithm does not spuriously detect outliers. Applying the modified algorithm of Shin et al. (1996) to AO contaminated series increases the size of the test back to its nominal significance level in all sample sizes independent of the outlier magnitude. In IO contaminated series the application of the algorithm does not influence the size properties. In fact, the size is not deteriorated by IOs, anyway.

Figures 5.8 and 5.9 present the power properties of the Leybourne test in the adjusted series. In the uncontaminated series the power is not affected by the adjustment of the series. The application of the modified algorithm of Shin et al. (1996) to AO contaminated series increases the power especially in series with large outlier magnitudes and equals the power in the uncontaminated series. In IO contaminated series the power increases and is higher than in the uncontaminated series. This is due to the fact that the algorithm can detect IOs only in the stationary part and thus, the differentiation between the stationary and the nonstationary part becomes easier (cf. Fig. 5.6).

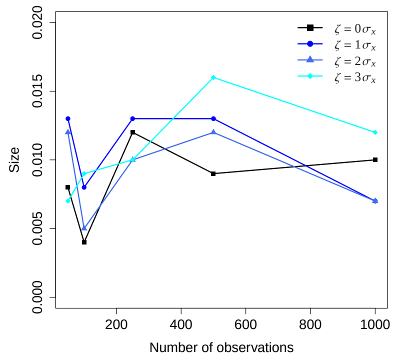


(a) Contaminated Series

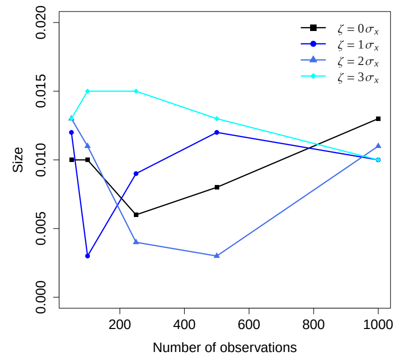


(b) Adjusted Series

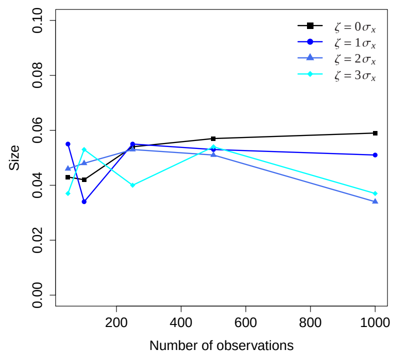
Figure 5.6: Influence of the Adjustment on an IO Contaminated Series with a Change in Persistence ($I(0) \rightarrow I(1)$)



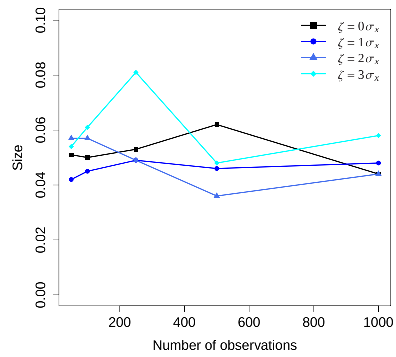
(a) AOs and $\alpha = 1\%$



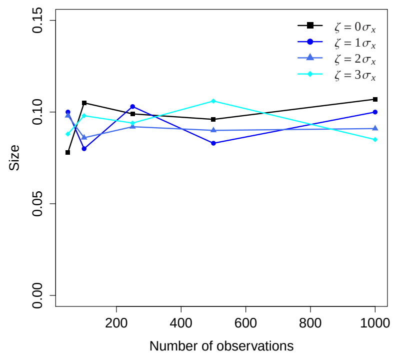
(b) IOs and $\alpha = 1\%$



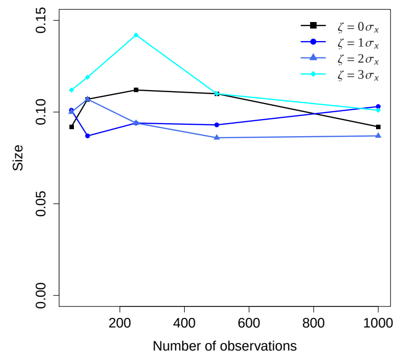
(c) AOs and $\alpha = 5\%$



(d) IOs and $\alpha = 5\%$

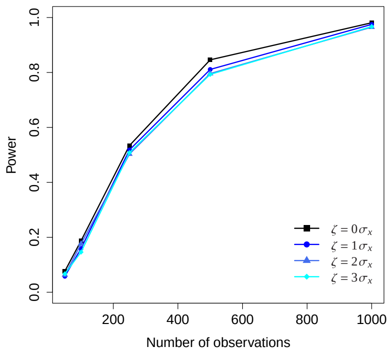


(e) AOs and $\alpha = 10\%$

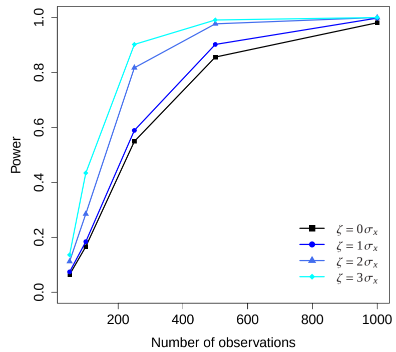


(f) IOs and $\alpha = 10\%$

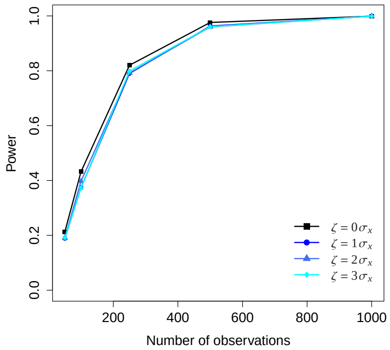
Figure 5.7: Size of the Leybourne Test ($I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance.



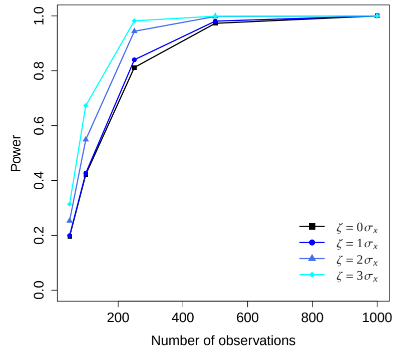
(a) AOs and $\alpha = 1\%$



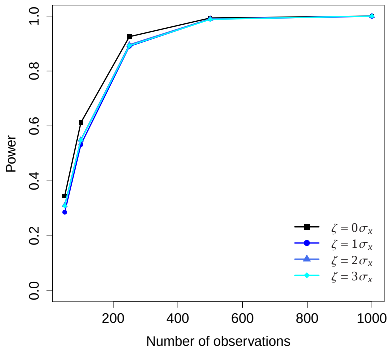
(b) IOs and $\alpha = 1\%$



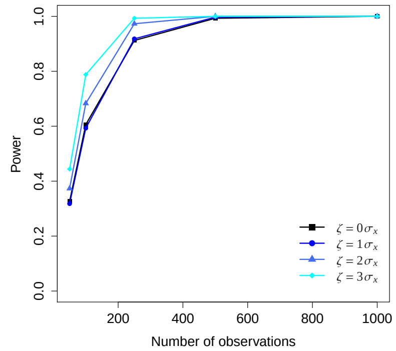
(c) AOs and $\alpha = 5\%$



(d) IOs and $\alpha = 5\%$

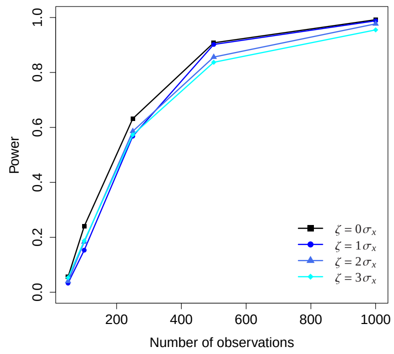


(e) AOs and $\alpha = 10\%$

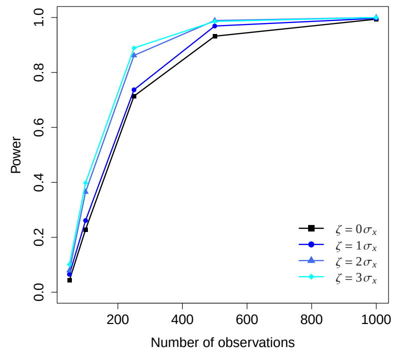


(f) IOs and $\alpha = 10\%$

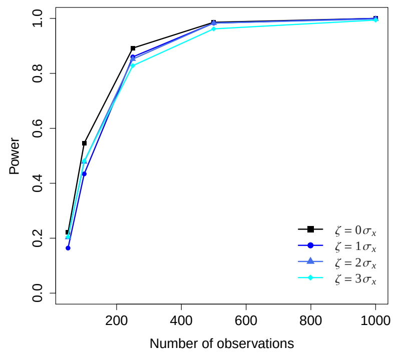
Figure 5.8: Power of the Leybourne Test ($I(0) \rightarrow I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance.



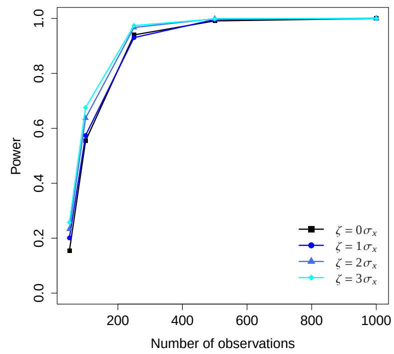
(a) AOs and $\alpha = 1\%$



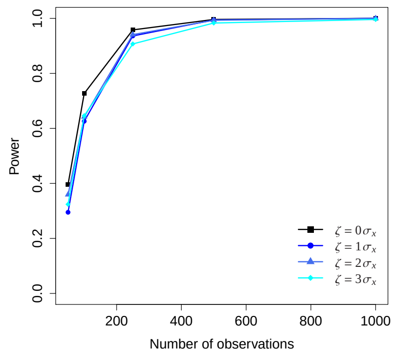
(b) IOs and $\alpha = 1\%$



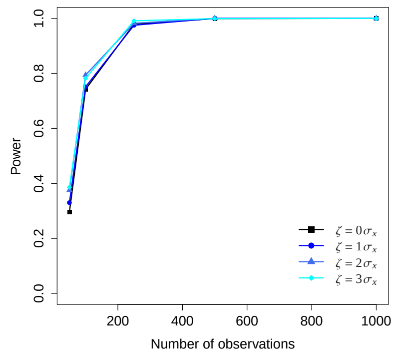
(c) AOs and $\alpha = 5\%$



(d) IOs and $\alpha = 5\%$



(e) AOs and $\alpha = 10\%$



(f) IOs and $\alpha = 10\%$

Figure 5.9: Power of the Leybourne Test ($I(1) \to I(0)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance.

6 Empirical Example

In this section we apply the tests for a change in persistence of Kim (2000); Kim et al. (2002) and of Leybourne et al. (2007) and the outlier detection method of Shin et al. (1996) to inflation data of the G7 countries. Following Busetti and Taylor (2004), we use quarterly CPI data from the OECD retrieved from FRED from 1970Q1 until 2014Q4 and define the inflation rates as

$$\pi_t = \log(CPI_t) - \log(CPI_{t-1}).$$

Thus, our data set consists of 180 observations for each country. We use the R package $X13$ for seasonal adjustment. The properties of the series change over time. During the Great Inflation in the 1970s and early 1980s inflation rates appear to exhibit a higher degree of persistence. At the beginning of the 1980s there is an overall decrease in the persistence. This period is referred to as the Great Moderation. The transition of the Great Inflation to the Great Moderation could present a change in persistence.

In Table 6.1 the critical values of both tests for a sample size of $T = 180$ are presented.

	Quantile					
	0.005	0.025	0.050	0.950	0.975	0.995
Kim	0.625	1.049	1.358	17.490	22.153	34.887
Leybourne	0.109	0.186	0.246	4.101	5.469	9.382

Table 6.1: Simulated Critical Values for $T = 180$

The test statistics of the Kim and Leybourne test applied to the G7 inflation rates for the original and the adjusted series are given in Table 6.2. Bold numbers indicate the rejection of the null hypothesis. In the original series the Kim test rejects the null hypothesis for Japan at the 10% significance level with an estimated change in 2005Q4. The Leybourne test rejects the null hypothesis in the original series for France at the 10% significance level and for the USA at the 1% level with the estimated changes in 1991Q4 and 1991Q1, respectively. After adjusting the series with the modified algorithm of Shin et al. (1996) the Kim test does not reject the null hypothesis for any country. In contrast, the Leybourne test rejects the null hypothesis for France at the 5% significance level and for Great Britain, Italy and Japan at the 10% significance level with the estimated changes in 1985Q2, 1990Q3, 1996Q1 and 1981Q4.

	CAN	FRA	GBR	GER	ITA	JPN	USA
Kim Test	4.0159	8.9347	3.0654	2.6954	4.5738	**1.1486**	6.9269
Leybourne Test	2.9018	**4.3994**	2.2575	1.7306	1.6157	2.7578	**23.0157**

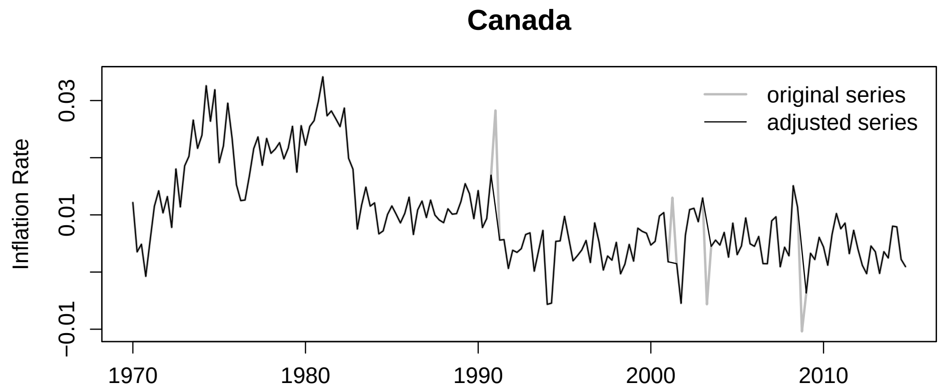
(a) Original Series

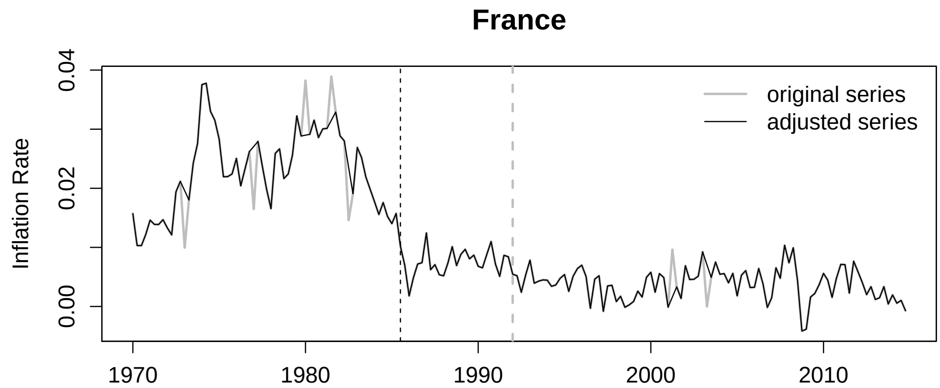
	CAN	FRA	GBR	GER	ITA	JPN	USA
Kim Test	3.7260	8.1539	2.3005	2.5883	6.6409	1.9316	5.8816
Leybourne Test	3.1757	**6.0435**	**5.0082**	1.9686	**4.9083**	**4.5627**	3.8323

(b) Adjusted Series

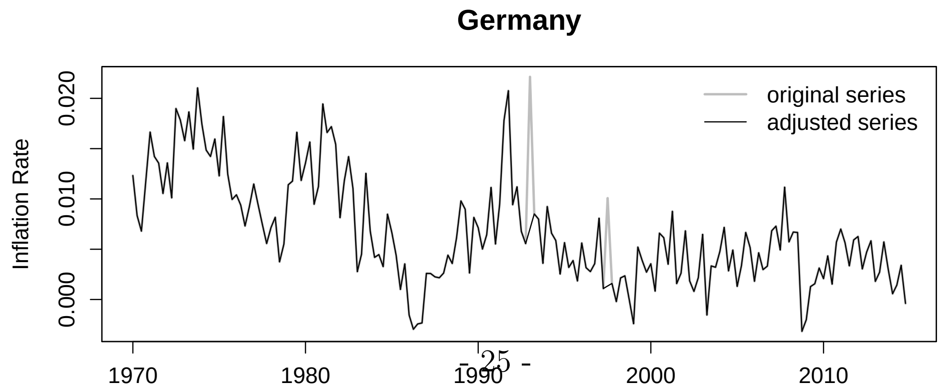
Table 6.2: Test Statistics of the Kim and Leybourne Test

In Figure 6.1 the original and the adjusted series of the G7 inflation rates are presented. The estimated change points are indicated by dashed lines.


Canada
Inflation Rate
−0.01
0.01
0.03
1970
1980
1990
2000
2010
original series
adjusted series


France
Inflation Rate
0.00
0.02
0.04
1970
1980
1990
2000
2010
original series
adjusted series


Great Britain
Inflation Rate
0.00
0.04
0.08
1970
1980
1990
2000
2010
original series
adjusted series


Germany
Inflation Rate
0.000
0.010
0.020
1970
1980
1990
2000
2010
original series
adjusted series

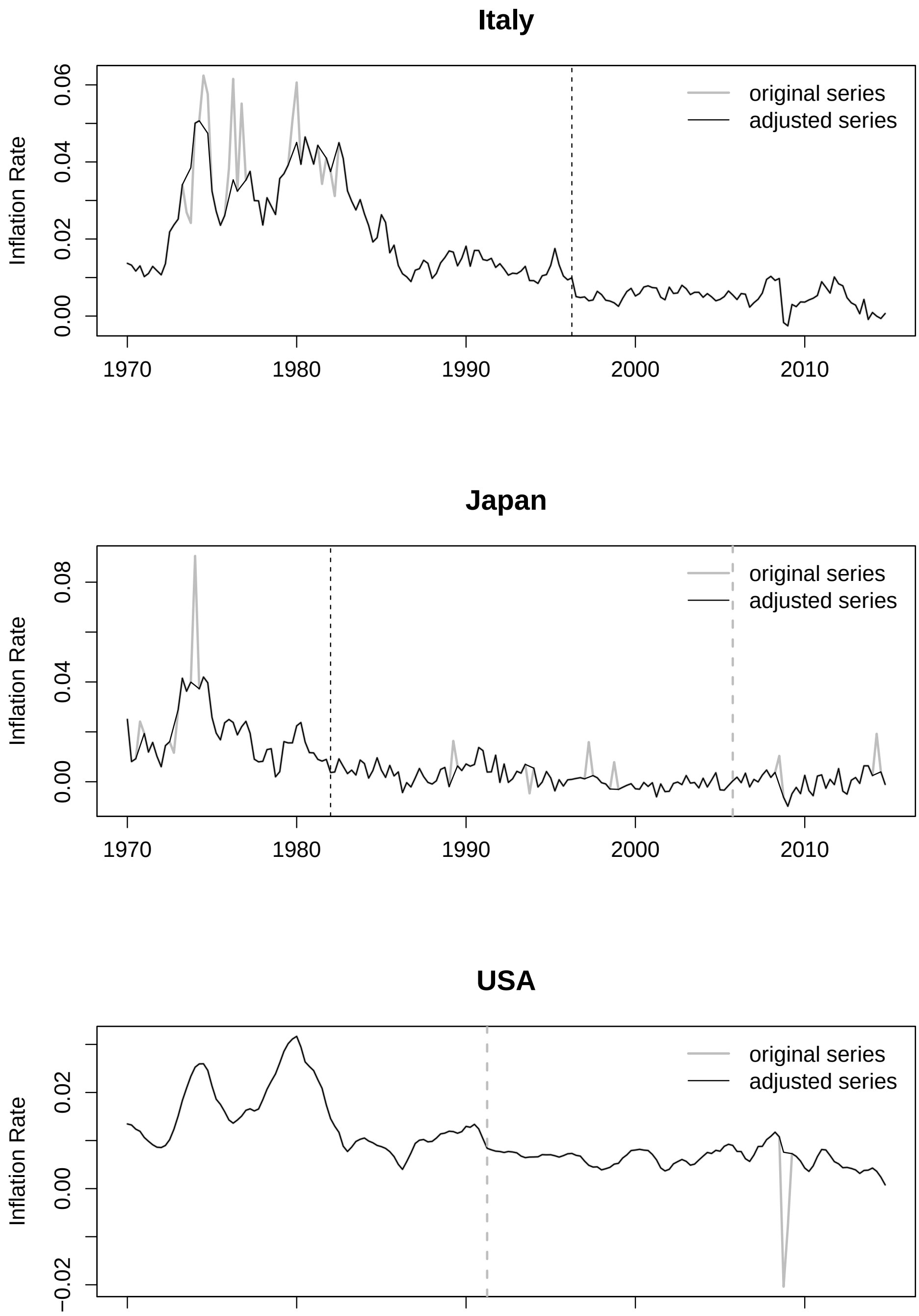


Figure 6.1: Inflation Rates of the G7 Countries

In order to support the test results, we conduct the unit-root test of Dickey and Fuller (1979).

Given the estimated change points the ADF test is conducted for the respective sub-samples. The p-values in Table 6.3 confirm the results of the tests for a change in persistence. Except for Japan in the original series and Italy the test detects a unit root in the first subsample and stationarity in the second subsample. This points to a change in persistence from $I(1)$ to $I(0)$ for France (in both series), the USA (in the original series), as well as Great Britain and Japan (in the adjusted series). Although the results for Italy are not as conclusive as for other countries, the p-values differ among the two subsamples. In the first subsample the null hypothesis can be rejected at the 10% level, whereas in the second subsample the p-value falls below 2%. Therefore we conclude that there occurs a change in persistence from $I(1)$ to $I(0)$ in the Italian series, which is also supported by the time series plot in Figure 6.1. In contrast, for Japan in the original series the p-values and the time series plot do not indicate a change in persistence. We deduce that the result of the Kim test is due to a type I error and that there is no change in persistence.

	FRA	JPN	USA
1st subsample	0.2625	0.0817	0.2176
2nd subsample	< 0.01	0.0980	< 0.01

(a) Original Series

	FRA	GBR	ITA	JPN
1st subsample	0.5386	0.2486	0.0741	0.3337
2nd subsample	0.0296	< 0.01	0.0195	0.0442

(b) Adjusted Series

Table 6.3: Subsample p-values of the ADF-Test

Summarizing our results we find different test decisions for the original and the adjusted series for four of the G7 countries. In Great Britain, Italy, and Japan the Leybourne test cannot detect a change in persistence in the original series due to outlier contaminations but confuses the series with a stationary process. After adjusting the series the Leybourne test rejects the null hypothesis in favor of a change in persistence from $I(1)$ to $I(0)$ which is supported by the results of the subsample ADF tests.

7 Conclusion

In this paper the effect of two different types of outliers on the performance of the tests for a change in persistence of Kim (2000); Kim et al. (2002) and of Leybourne et al. (2007) are assessed. We find that the Kim test is not seriously affected by outliers. Especially the size of the test is not deteriorated. Due to the low degree of persistence under the null hypothesis of the test, AOs and IOs have the same effect on the series under the null hypothesis. The contaminated stationary process is identified as a stationary process and thus the size is not affected. Therefore, we conclude that it is not necessary to detect and remove outliers before applying the test. In contrast, the Leybourne test suffers from severe size and power distortions due to AOs. IOs do not affect the size but can even lead to power gains. As a result, we recommend to adjust the contaminated series and remove AOs before applying the test. The modified algorithm of Shin et al. (1996) performs well and is easy to implement. After adjusting the series, the size of the test coincides with the nominal significance levels and the power converges to 1 with an increasing sample size. In the empirical application we use the tests to find changes in persistence in the G7 inflation rates. We detect a change in persistence for France in the original and the adjusted series, and for Great Britain, Italy, and Japan after adjusting the series.

A Appendix

A.1 Limiting Distribution

Suppose the core process of the data generating process is a random walk, which coincides with the null hypothesis of the Leybourne test,

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

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. The observable series $\{y_t\}$ is contaminated with AOs of magnitude ζ if $\delta_t = \pm 1$,

$$y_t = x_t + \zeta \delta_t.$$

For an AO at $t = s$, we obtain

$$y_{s-1} = x_{s-1} = x_{s-2} + \varepsilon_{s-1} \qquad y_s = x_s + \zeta = x_{s-1} + \varepsilon_s + \zeta \qquad y_{s+1} = x_{s+1} = x_s + \varepsilon_{s+1}.$$

Under the assumption of $\{y_t\}$ being a random walk, the residuals are given by

$$e_s = \varepsilon_s + \zeta \qquad e_{s+1} = \varepsilon_{s+1} - \zeta,$$

where $e_s \sim N(\zeta, \sigma_\varepsilon^2)$ and $e_{s+1} \sim N(-\zeta, \sigma_\varepsilon^2)$ (cf. Shin et al., 1996). The linear combination $e_{s+1} - e_s$ follows a normal distribution with $\mu = -2\zeta$ and $\sigma^2 = 2\sigma_\varepsilon^2$. If the random variable $X \sim N(\mu, \sigma^2)$, then $Z = |X|$ follows a folded normal distribution (cf. Leone et al., 1961), with the density function

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \left[\exp\left(-\frac{(z-\mu)}{2\sigma^2}\right) + \exp\left(-\frac{(z+\mu)}{2\sigma^2}\right) \right].$$

Under the null hypothesis the series $\{y_t\}$ is uncontaminated and therefore $\zeta = 0$. Thus, $e_s, e_{s+1} \sim N(0, \sigma_\varepsilon^2)$, $e_{s+1} - e_s \sim N(0, 2\sigma_\varepsilon^2)$, and $|e_{s+1} - e_s|$ follows a folded normal distribution with density function

$$f(|e_{s+1} - e_s|) = \frac{1}{\sqrt{\pi\sigma_\varepsilon^2}} \exp\left(-\frac{(|e_{s+1} - e_s|)^2}{4\sigma_\varepsilon^2}\right).$$

This coincides with twice the right tail of the normal distribution $N(0, 2\sigma_\varepsilon^2)$.
The test statistic

$$\lambda^* = \frac{1}{\sqrt{2\hat{\sigma}_\varepsilon^2}} (|e_{t+1} - e_t|) \qquad t = 2, \ldots, (T-1),$$

where $\hat{\sigma}_\varepsilon^2$ is a robust estimator for the error term variance σ_ε^2, hence follows a standard folded normal distribution. Critical values can be obtained according to $q_{1-\alpha}^{\lambda^*} = z_{1-\alpha/2}$ for $\alpha \leq 0.5$, where z is a quantile of the standard normal distribution.
Figure A.1 and Table A.1 illustrate the convergence of the test statistic λ^* to the standard folded normal distribution.

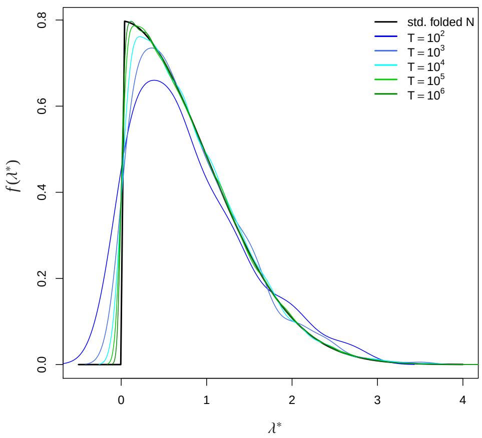


Figure A.1: Estimated Density of λ^* and the Standard Folded Normal Distribution

	0.01	0.05	0.1	0.9	0.95	0.99
T = 10^2	0.0109	0.0412	0.0704	1.8610	2.2003	2.7312
T = 10^3	0.0109	0.0641	0.1323	1.6383	2.0642	2.6630
T = 10^4	0.0146	0.0681	0.1294	1.6431	1.9664	2.6194
T = 10^5	0.0129	0.0613	0.1242	1.6518	1.9667	2.5872
T = 10^6	0.0124	0.0629	0.1256	1.6448	1.9623	2.5773
std. folded N	0.0125	0.0627	0.1257	1.6448	1.9600	2.5758

Table A.1: Quantiles of the Estimated Density of λ^* and of the Standard Folded Normal Distribution

Since it is not known a priori when an AO occurs, the maximum of the absolute difference between two consecutive residuals is taken,

$$\lambda = \frac{1}{\sqrt{2\hat{\sigma}_\varepsilon^2}}\left(\max_{2\leq t\leq (T-1)} |e_{t+1} - e_t|\right).$$

According to the extreme value theory, the maximum of random variables from a distribution of the exponential family follows the Gumbel distribution (cf. Gumbel, 1958, pp. 164f, Kotz and Nadarajah, 2000, p. 59) with density function

$$f(x) = \frac{1}{\beta}\exp\left(-\frac{x-\mu}{\beta} + \exp\left(-\frac{x-\mu}{\beta}\right)\right).$$

However, the test statistic λ does not follow a standard Gumbel distribution ($\mu = 0$ and $\beta = 1$) since the standard Gumbel distribution allows for negative realizations (cf. Tab. A.2), whereas the absolute does not. In order to determine appropriate critical values, we calculate λ for random walks of different sample sizes $T = \{10^2, 10^3, 10^4, 10^5\}$, each with 1000 replications. We find that the distribution of λ crucially depends on the sample size. For an increasing number of observations, the distribution shifts to the right and the quantiles increase (cf. Tab. A.2).

	0.01	0.05	0.1	0.9	0.95	0.99
T = 10^2	2.1052	2.2251	2.3351	3.3154	3.5183	4.0632
T = 10^3	2.8462	2.9854	3.0545	3.8856	4.0141	4.3855
T = 10^4	3.4985	3.6176	3.6831	4.4230	4.6072	4.8394
T = 10^5	4.0520	4.1789	4.2331	4.8525	4.9963	5.3274
std. Gumbel	-1.5272	-1.0972	-0.8340	2.2504	2.9702	4.6001

Table A.2: Quantiles of the Estimated Density of λ and of the Standard Gumbel Distribution

In addition to the sample size, the distribution of λ also depends on the number of iterations. If the test is applied more than once to the (adjusted) series, the distribution shifts to the left. For an increasing number of iterations (detection of the maximum in the adjusted series), the distribution asymptotically converges to the standard folded normal distribution. The estimated quantiles of λ based on 10000 replications for different sample sizes $T = \{100, 500, 1000\}$ and different numbers of iterations $\{1, 2, 9, 100\}$ are illustrated in Figure A.2.

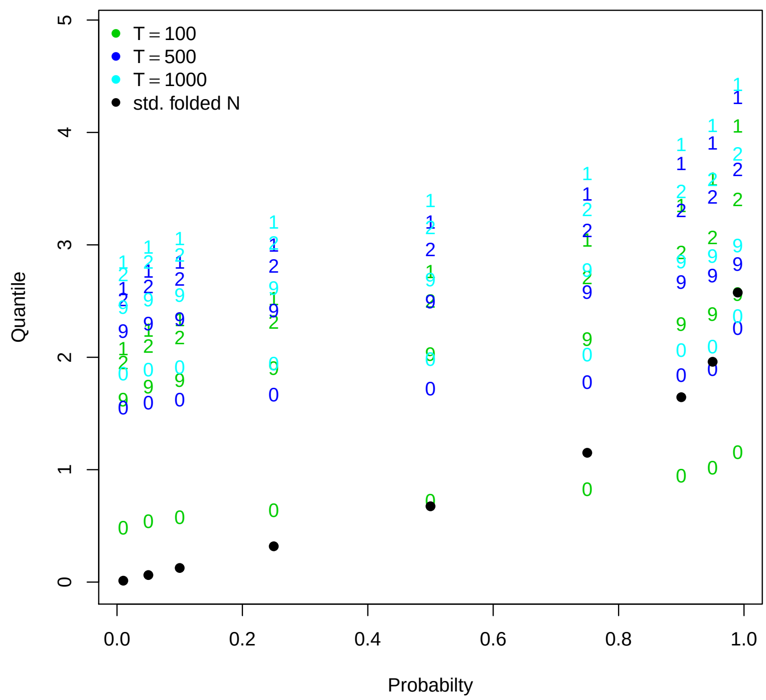
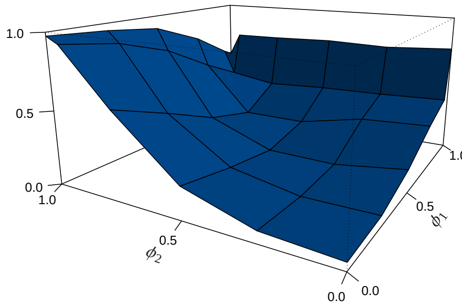


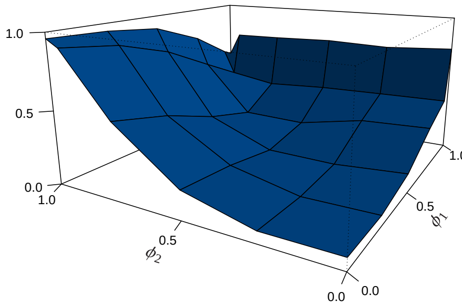
Figure A.2: Estimated Quantiles of λ after Different Iterations

From Figure A.2 we conclude that a critical value for λ cannot be derived from a limiting distribution since it is not clear beforehand how many iterations are needed to remove AOs from the series. Applying a large critical value reduces the risk of falsely identifying outliers, but may prevent the algorithm from detecting true outliers. In contrast, using a small critical value guarantees that outliers are correctly identified, but will also lead to spurious detection of outliers. The critical value of $C = 3$ recommended by Shin et al. (1996) seems to balance this trade-off. On the one hand the probability for a standard folded normal distributed random variable to exceed a value of $C = 3$ only amounts to 0.270%. Therefore, we do not expect the algorithm to detect many falsely classified outliers or to get stuck in an endless loop. On the other hand according to Figure A.2 the critical value of $C = 3$ is small enough for the test not to be conservative.

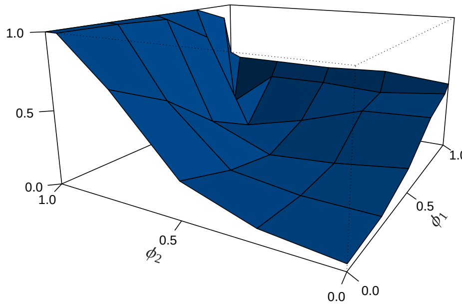
A.2 Power Plots



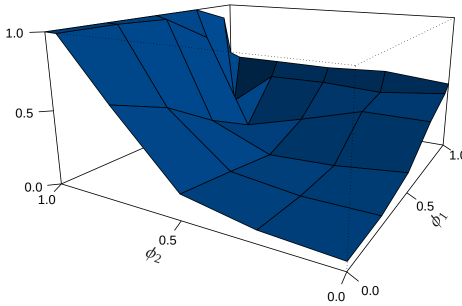
(a) Original Series, $T = 100$, $\zeta = 0$



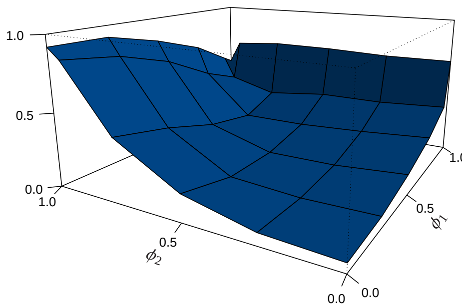
(b) Adjusted Series, $T = 100$, $\zeta = 0$



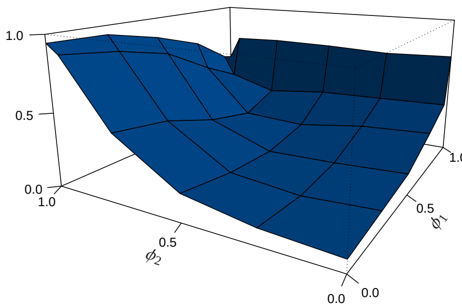
(c) Original Series, $T = 1000$, $\zeta = 0$



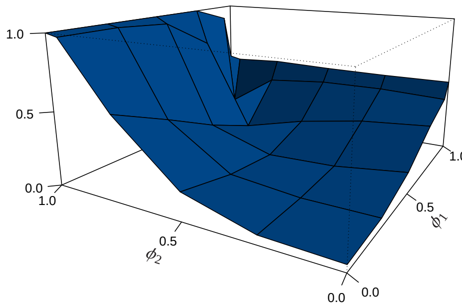
(d) Adjusted Series, $T = 1000$, $\zeta = 0$



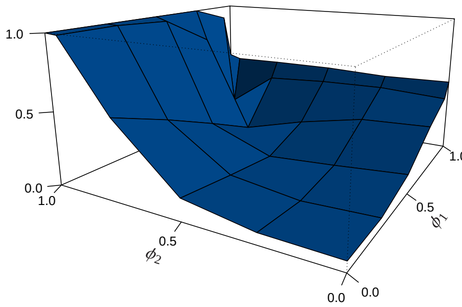
(e) Original Series, $T = 100$, $\zeta = 3$



(f) Adjusted Series, $T = 100$, $\zeta = 3$

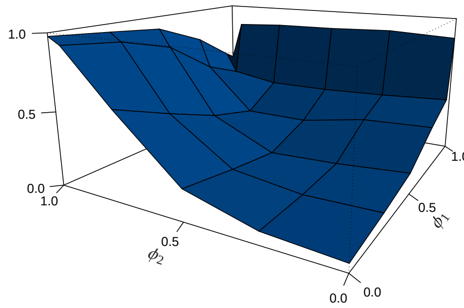


(g) Original Series, $T = 1000$, $\zeta = 3$

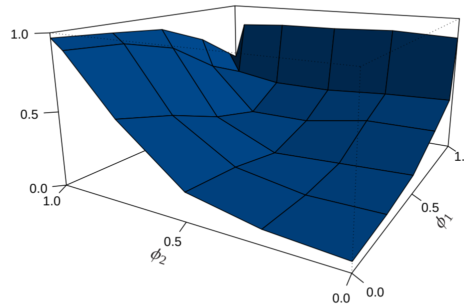


(h) Adjusted Series, $T = 1000$, $\zeta = 3$

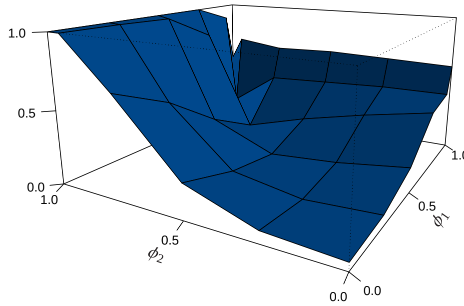
Figure A.3: Power of the Kim Test for Additive Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.



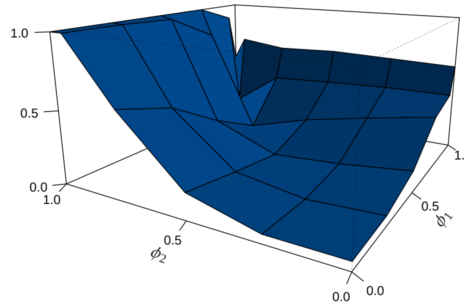
(a) Original Series, $T = 100$, $\zeta = 0$



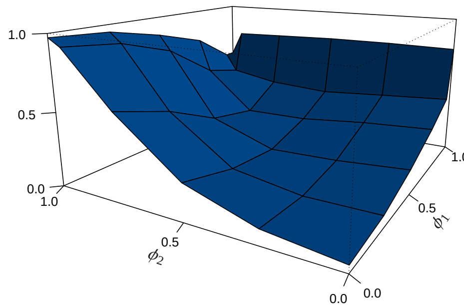
(b) Adjusted Series, $T = 100$, $\zeta = 0$



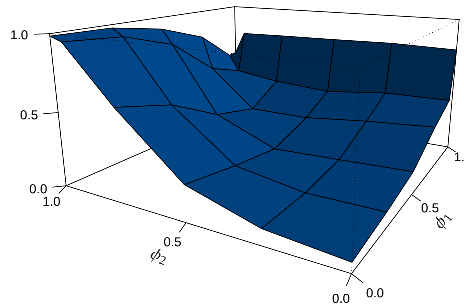
(c) Original Series, $T = 1000$, $\zeta = 0$



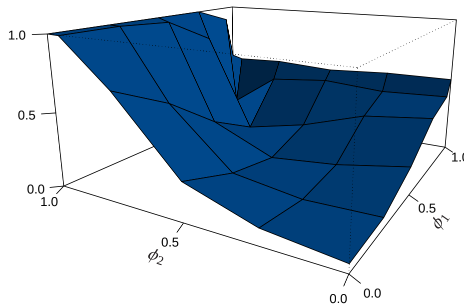
(d) Adjusted Series, $T = 1000$, $\zeta = 0$



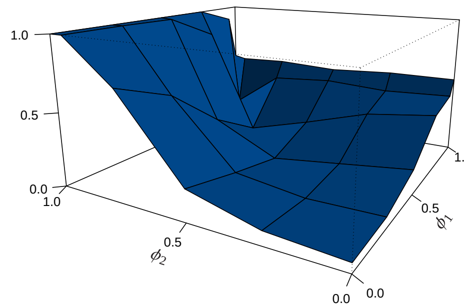
(e) Original Series, $T = 100$, $\zeta = 3$



(f) Adjusted Series, $T = 100$, $\zeta = 3$

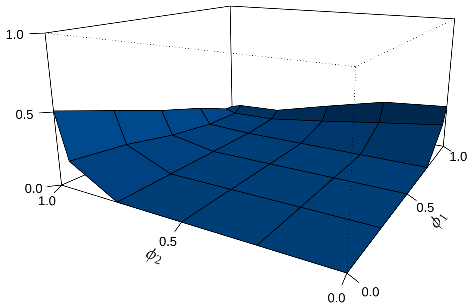


(g) Original Series, $T = 1000$, $\zeta = 3$

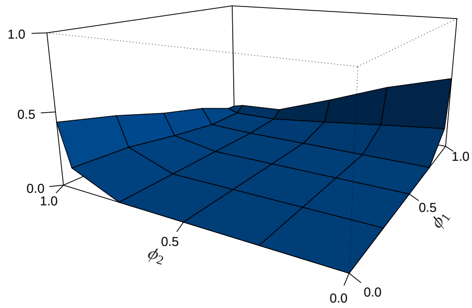


(h) Adjusted Series, $T = 1000$, $\zeta = 3$

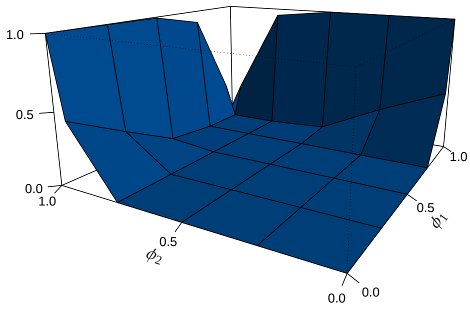
Figure A.4: Power of the Kim Test for Innovative Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.



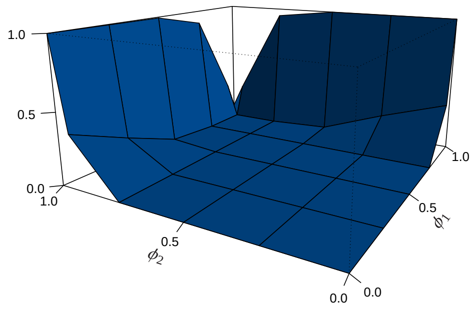
(a) Original Series, $T = 100$, $\zeta = 0$



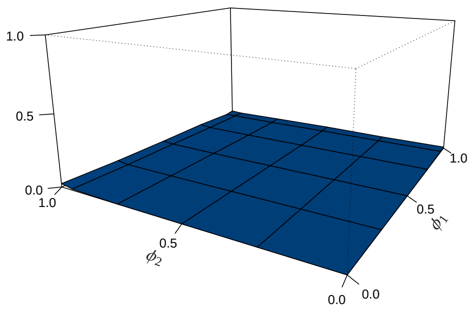
(b) Adjusted Series, $T = 100$, $\zeta = 0$



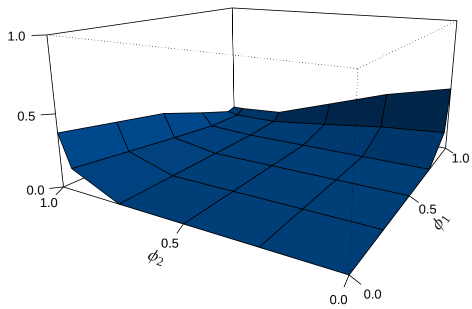
(c) Original Series, $T = 1000$, $\zeta = 0$



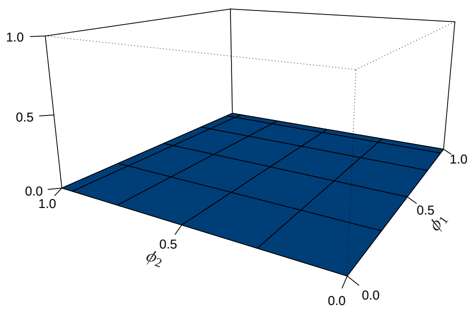
(d) Adjusted Series, $T = 1000$, $\zeta = 0$



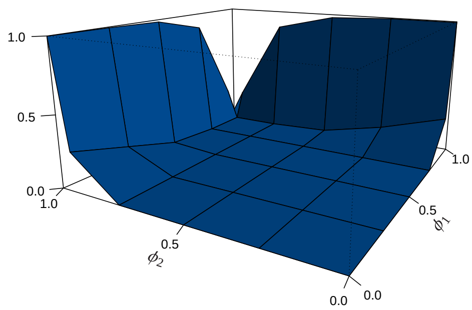
(e) Original Series, $T = 100$, $\zeta = 3$



(f) Adjusted Series, $T = 100$, $\zeta = 3$

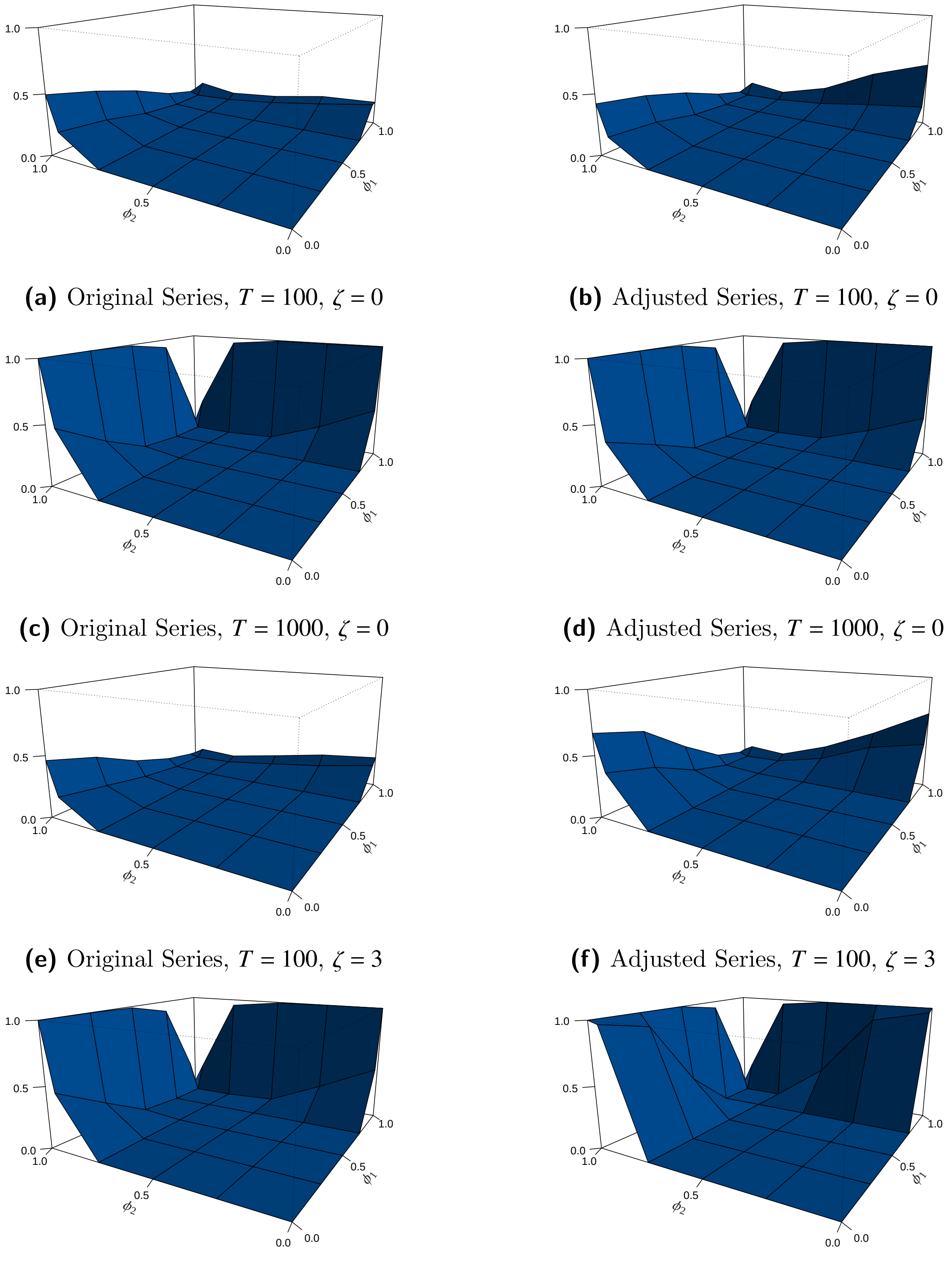


(g) Original Series, $T = 1000$, $\zeta = 3$



(h) Adjusted Series, $T = 1000$, $\zeta = 3$

Figure A.5: Power of the Leybourne Test for Additive Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.



(a) Original Series, $T = 100$, $\zeta = 0$

(b) Adjusted Series, $T = 100$, $\zeta = 0$

(c) Original Series, $T = 1000$, $\zeta = 0$

(d) Adjusted Series, $T = 1000$, $\zeta = 0$

(e) Original Series, $T = 100$, $\zeta = 3$

(f) Adjusted Series, $T = 100$, $\zeta = 3$

(g) Original Series, $T = 1000$, $\zeta = 3$

(h) Adjusted Series, $T = 1000$, $\zeta = 3$

Figure A.6: Power of the Leybourne Test for Innovative Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.

References

Ahmad, Y. and Donayre, L. (2016). Outliers and persistence in threshold autoregressive processes. *Studies in Nonlinear Dynamics & Econometrics*, 20(1):37–56.

Busetti, F. and Taylor, A. R. (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics*, 123(1):33–66.

Chang, I., Tiao, G. C., and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193–204.

Chen, C. and Liu, L.-M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421):284–297.

Davies, L. and Gather, U. (1993). The identification of multiple outliers. *Journal of the American Statistical Association*, 88(423):782–792.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(3):350–363.

Franses, P. H. and Haldrup, N. (1994). The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business & Economic Statistics*, 12(4):471–478.

Galeano, P. and Peña, D. (2013). Finding outliers in linear and nonlinear time series. In *Robustness and Complex Data Structures*, pages 243–260. Springer.

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430.

Hansen, B. E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics & Econometrics*, 2(1).

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, pages 1551–1580.

Kim, J.-Y. (2000). Detection of change in persistence of a linear time series. *Journal of Econometrics*, 95(1):97–116.

Kim, J.-Y., Belaire-Franch, J., and Amador, R. B. (2002). Corrigendum to "Detection of change in persistence of a linear time series" [J. Econom. 95 (2000) 97–116]. *Journal of Econometrics*, 109(2):389–392.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press.

Leone, F. C., Nelson, L. S., and Nottingham, R. B. (1961). The folded normal distribution. *Technometrics*, 3(4):543–550.

Leybourne, S., Kim, T.-H., Smith, V., and Newbold, P. (2003). Tests for a change in persistence against the null of difference-stationarity. *The Econometrics Journal*, 6(2):291–311.

Leybourne, S., Taylor, R., and Kim, T.-H. (2004). An unbiased test for a change in persistence. *Department of Economics Discussion Paper-University of Birmingham*.

Leybourne, S., Taylor, R., and Kim, T.-H. (2007). Cusum of squares-based tests for a change in persistence. *Journal of Time Series Analysis*, 28(3):408–433.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499.

Martin, R. D. and Yohai, V. J. (1986). Influence functionals for time series. *The Annals of Statistics*, 14(3):781–818.

Peña, D. (1991). Measuring influence in dynamic regression models. *Technometrics*, 33(1):93–101.

Sánchez, M. J. and Peña, D. (2003). The identification of multiple outliers in ARIMA models. *Communications in Statistics-Theory and Methods*, 32(6):1265–1287.

Shin, D. W., Sarkar, S., and Lee, J. H. (1996). Unit root tests for time series with outliers. *Statistics & Probability Letters*, 30(3):189–197.

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7(1):1–20.

van Dijk, D., Franses, P. H., and Lucas, A. (1999). Testing for smooth transition nonlinearity in the presence of outliers. *Journal of Business & Economic Statistics*, 17(2):217–235.

van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002). Smooth transition autoregressive models - A survey of recent developments. *Econometric Reviews*, 21(1):1–47.