Essays on Nonlinearities in Time Series: Regime Switching, Outlying Observations, and Changes in Persistence

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
- Doctor rerum politicarum -

genehmigte Dissertation

von

M.Sc. Saskia Rinke

geboren am 13. Juni 1989 in Bad Oldesloe

2017

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Koreferentin: Prof. Dr. Ulrike Grote, Leibniz Universität Hannover

Tag der Promotion: 20.12.2016

Acknowledgements

First of all, I would like to thank my advisor and co-author Philipp Sibbertsen for giving me the opportunity to earn my PhD degree at the Institute of Statistics and for his support as well as his advise in the last three years.

Furthermore, I am indebted to Ulrike Grote for agreeing to be the second examiner of my thesis and to Andreas Wagener and Philip Bertram for joining my examination board.

I am also grateful to my other co-author Tristan Hirsch and to my other colleagues for creating such a friendly and productive working environment.

Special thanks goes to my family and friends for their permanent support, encouragement, and confidence.

Saskia Rinke
Hannover, September 2016

Abstract

Economic and financial time series often exhibit nonlinear features. In order to be able to correctly identify the effect of shocks and to generate reliable forecasts, an appropriate model has to be fitted to the data. In doing so, the trade-off between goodness of fit and model complexity becomes apparent: Are the nonlinear features of the data important for further inference or can a simpler model be applied?
In this thesis the model selection procedure between linear and nonlinear time series models is studied with a focus on regime-switching models and structural breaks.

In Chapter 2 the performance of information criteria to detect nonlinearity and to distinguish between different transition functions is assessed in an extensive simulation study. Since information criteria perform well in linear models, different versions are suggested that to some degree exploit the linearity of the single regimes of the nonlinear models. Certain information criteria and specific versions are recommended depending on the aim of application.

In Chapter 3 the influence of outlying observations on the model selection procedure is investigated. Outliers generate spurious nonlinearity in linear time series. Linearity tests for example suffer from severe size distortions in the presence of outliers. In several simulation studies the sensitivity of information criteria to additive outliers is assessed. The results suggest that information criteria are more robust to additive outliers than linearity tests.

Chapter 4 deals with the effect of outliers on the performance of the tests for a change in persistence. In the simulation study the performance of the tests in uncontaminated, outlier contaminated, and adjusted series is investigated and compared. Since additive outliers deteriorate the performance of the tests if the series under consideration exhibits a high degree of persistence, additive outliers have to be located and removed before the application of the tests. A modified outlier detection and removal method from the unit-root literature is recommended. The tests and the outlier detection algorithm are applied to inflation data of the G7 countries.

Keywords: Information Criteria · Nonlinearity · Additive Outliers · Innovative Outliers · Change in Persistence · Outlier Detection

Zusammenfassung

Ökonomische und Finanzzeitreihen weisen oftmals nichtlineare Charakteristika auf. Damit die weitere Inferenz zuverlässige Ergebnisse liefert, muss ein Modell an die vorliegenden Daten angepasst werden, das sämtliche wichtige Eigenschaften der Daten abbilden kann. Dabei tritt der bekannte Zielfkonflikt zwischen Güte der Anpassung und Modellkomplexität auf: Es gilt zu entscheiden, ob die nichtlinearen Eigenschaften der Daten wichtig für die weitere Analyse sind und in das Modell mitaufgenommen werden oder ob ein einfacheres Modell ausreichend ist.
Die vorliegende Arbeit behandelt die Modellselektion zwischen linearen und nichtlinearen Zeitreihenmodellen mit einem Fokus auf Regimewechsel-Modellen und Strukturbrüchen.

In Kapitel 2 wird mittels diverser Simulationsstudien untersucht, inwieweit Informationskriterien genutzt werden können, um Nichtlinearität zu erkennen und die Form der Übergangsfunktion nichtlinearer Modelle zu bestimmen. Da Informationskriterien sehr gut geeignet sind, um zwischen linearen Modellen zu selektieren, werden verschiedene Versionen der Informationskriterien vorgeschlagen, die zum Teil die Linearität der einzelnen Regime ausnutzen. Je nach Ziel der Anwendung werden bestimmte Versionen einzelner Informationskriterien empfohlen.

In Kapitel 3 wird der Einfluss von Ausreißern auf die Modellselektion untersucht. Ausreißer erzeugen scheinbare Nichtlinearität in linearen Prozessen. Daher steigt z.B. die Wahrscheinlichkeit für den Fehler erster Art bei Linearitätstests stark an. In umfangreichen Simulationsstudien wird die Sensitivität von Informationskriterien gegenüber Ausreißern untersucht. Die Ergebnisse zeigen, dass Informationskriterien robuster gegenüber additiven Ausreißern sind als Linearitätstests.

Kapitel 4 behandelt den Einfluss von Ausreißern auf Persistenzbruchtests. In verschiedenen Simulationsstudien wird das Abschneiden der Tests in unkontaminierten, kontaminierten und bereinigten Reihen untersucht und verglichen. Die Ergebnisse zeigen, dass additive Ausreißer in Reihen mit einer hohen Persistenz die Güte der Tests verzerren. Additive Ausreißer müssen demnach vorab lokalisiert und die Reihe um deren Effekte bereinigt werden. Ein modifizierter Algorithmus aus der Einheitswurzel-Literatur wird dafür empfohlen. Die Tests und der Bereinigungsalgorithmus werden auf Inflationsdaten der G7 Länder angewendet.

Stichworte: Informationskriterien · Nichtlinearität · Additive Ausreißer · Innovative Ausreißer · Persistenzbruch · Ausreißerermittlung

Contents

Chapter 1

Introduction

Introduction

For several decades linear models dominated in time series analysis and econometrics, starting with the autoregressive model of Yule (1927) and becoming more frequently used after the introduction of the modeling approach of Box and Jenkins (1970). This is due to the fact that linear models allow for a good model fit with a parsimonious parametrization, they are easy to estimate, and according to Tong and Lim (1980) they yield good one-step-ahead predictions. Despite these desirable properties, linear models suffer from serious shortcomings when applied to economic or financial time series. Real data often exhibits nonlinear features, like regime-switching behavior or a structural break. Examples for these types of nonlinearity include the asymmetry between recessions and expansions in the business cycle (cf. van Dijk et al., 2002) and the change in the degree of persistence in inflation rates due to monetary policy changes (cf. Bai and Perron, 2003 among others). Since the existing class of linear models was not able to capture these stylized facts, multiple-regime models as well as tests and estimation procedures to detect different types of structural breaks were introduced to incorporate nonlinearity into time series models. As a result, the model fit and the forecast performance are improved, but also the complexity of the models increases compared to linear models. So, there exists the well-known trade-off between goodness of fit and model complexity.

In this thesis the model selection procedure between linear and nonlinear time series models is studied in a univariate framework. It consists of three independent chapters dealing with different aspects of the model selection between linear and nonlinear models. The main focus is on regime-switching models in Chapters 2 and 3 and on changes in persistence in Chapter 4.

In Chapter 2, co-authored with Philipp Sibbertsen, the model selection between linear autoregressive and nonlinear threshold autoregressive models using information criteria is considered. Regime-switching models were introduced by Tong and Lim (1980), Chan and Tong (1986), and Hamilton (1989). The idea of this model class is to extend the linear autoregressive model with its desirable properties to capture nonlinearity. In threshold autoregressive models two or more linear regimes are linked via a transition function and are combined into one model. A switch between the regimes occurs if a fixed threshold value is exceeded. In case of the self-exciting threshold autoregressive model (SETAR) of Tong and Lim (1980) and Tong (1983), the transition function is the indicator function, implying that the shifts between different regimes are discreet jumps. Replacing the indicator function with a continuous function that takes values in the unit

interval, the smooth transition autoregressive model (STAR) of Chan and Tong (1986) (cf. also Teräsvirta, 1994) is obtained. Depending on the form of the transition function, STAR models can be symmetric like the ESTAR or asymmetric like the LSTAR model.
To select among different linear models, information criteria are frequently applied. However, it is not clear whether the criteria maintain their optimality properties if they are applied to nonlinear models (cf. Clements and Krolzig, 1998) and thus, can be used instead of linearity tests to detect nonlinearity. Therefore, in Chapter 2 different versions of information criteria are proposed which to some extent exploit the linearity of the single regimes of the threshold autoregressive models. Their performance is assessed in an extensive simulation study. The results suggest that the information criteria perform well in general. Several factors influencing the performance of the information criteria are identified and depending on the aim of application the use of certain versions and of particular information criteria is recommended.

In Chapter 3 the sensitivity of information criteria to outlying observations is assessed.
The nonlinear features of a process can be the result of only a few observations, especially in small to moderate samples (cf. van Dijk et al., 1999). In this case it is not clear whether nonlinearity is an important feature of the data that should be appropriately captured by a model or whether the few observations that generate the nonlinearity are outlying observations and the process exhibits only spurious nonlinearity. In the presence of outliers the discrimination between linear and nonlinear models using linearity tests becomes unreliable since the tests become oversized and spuriously reject the null hypothesis of linearity (cf. van Dijk et al., 1999 and Ahmad and Donayre, 2016). This is due to the fact that regime-switching models can generate data that resembles an outlier contaminated linear process (cf. van Dijk et al., 2002).
In Chapter 3 the performance of information criteria to detect nonlinearity in outlier contaminated series is compared to the linearity test against SETAR nonlinearity of Hansen (1999). The simulation results show that the information criteria can clearly outperform the linearity test in terms of size in larger samples and for larger outlier magnitudes. Hence, information criteria are able to detect spurious nonlinearity due to outlying observations more reliably.

In Chapter 4, co-authored with Tristan Hirsch, tests for a change in persistence are considered. In addition to regime-switching behavior, nonlinearity can also be induced by a structural break. This implies a change in at least one parameter value, i.e. the mean, the persistence, the trend or the variance. Examples of a structural break might be a lower growth rate after the oil price shock in 1973 (cf. Perron, 1989) or a change in the persistence of inflation rates due to different monetary policy regimes (cf. Bai and Perron, 2003 among others). Identifying a structural break is crucial since the data generating process changes and thus, the model structure has to be modified and updated to be able to appropriately capture the important features of the data, to correctly forecast the series and to be able to assess the true influence of shocks. Depending on the type of structural break, different tests and detection methods are available. An overview is given in Perron (2006).

The tests under consideration in Chapter 4 are designed to detect a structural break in the persistence. They exhibit a linear process under the null hypothesis, whereas under the alternative a change in persistence occurs and the series becomes nonlinear. In a simulation study the effect of outlying observations on the performance of the tests is assessed. The results show that there occur serious size and power distortions if the process is highly persistent. Therefore, a modified version of the outlier detection method of Shin et al. (1996) is applied to remove additive outliers before the application of the tests. In an empirical example, the tests and the detection method are applied to the quarterly inflation rates of the G7 countries.

Chapter 2

Information Criteria for Nonlinear Time Series Models

Information Criteria for Nonlinear Time Series Models

Co-authored with Philipp Sibbertsen

Published in Studies in Nonlinear Dynamics and Econometrics (2016) **20(3)**, *pp. 325—341*

Online available at: `https://doi.org/10.1515/snde-2015-0026`

Chapter 3

The Influence of Additive Outliers on the Performance of Information Criteria to Detect Nonlinearity

The Influence of Additive Outliers on the Performance of Information Criteria to Detect Nonlinearity

3.1 Introduction

Model identification is one of the major challenges in time series analysis due to the trade-off between goodness of fit and model complexity. In general, linear models perform as benchmark models because they yield a good model fit while being simple and easy to estimate. Although nonlinear models are more complex, they may be able to better capture certain characteristics of the true data generating process (DGP), which improve the forecasting performance enormously. However, there are factors like outlying observations that further complicate the model selection procedure.

In order to select between linear and nonlinear models, generally, linearity tests are conducted. Outliers can lead to serious size distortions of linearity tests and hence to a spurious selection of nonlinear models. This is due to the fact that multiple regime models can generate data resembling an outlier contaminated linear process (cf. van Dijk et al., 2002). So, van Dijk et al. (1999) find that the test against smooth transition nonlinearity of Luukkonen et al. (1988b) becomes oversized in the presence of additive outliers (AOs) and thus, the null hypothesis of linearity is rejected too often. For a high outlier probability or large outlier magnitudes the size decreases again, but the power of the test is deteriorated. Recently, Ahmad and Donayre (2016) also detect size distortions but power improvements due to outliers for the test against threshold autoregressive nonlinearity of Hansen (1996, 1997).

As an alternative to linearity tests, information criteria can be used to detect nonlinearity. So, the application of information criteria to identify the number of regimes of autoregressive (AR) and self-exciting threshold autoregressive (SETAR) models is treated in Gonzalo and Pitarakis (2002), Hamaker (2009), and Rinke and Sibbertsen (2016). The selection of the model class in general with information criteria is also considered in Kapetanios (2001) and Psaradakis et al. (2009). Since the application of information criteria is less popular to detect nonlinearity, the performance of information criteria in outlier contaminated time series has not been assessed yet.

Therefore, in this paper the effect of AOs on model selection between AR and SETAR models using information criteria is investigated by means of simulations. The influence of the outlier magnitude, the degree of persistence, and the sample size on the performance of information criteria is assessed. Their performance is then compared to the performance of the linearity test against a SETAR alternative of Hansen (1999) in order to evaluate whether information criteria can be a useful alternative for model selection in outlier contaminated time series.

The rest of the paper is organized as follows. In Section 3.2 the model framework is presented. In Section 3.3 the methods to detect nonlinearity are introduced. Section 3.4 contains the simulation set-up and the simulation results. Finally, Section 3.5 concludes.

3.2 Additive Outliers in Linear and Nonlinear Time Series

According to Davies and Gather (1993) and van Dijk et al. (1999) outliers can only be defined in the context of a certain model. In this paper linear AR(1) and two-regime SETAR models consisting of AR(1) specifications in both regimes (SETAR(1,1)) are considered. The AR(1) model is defined as

$$x_t = \phi x_{t-1} + \varepsilon_t,$$

for $t = 1, \ldots, n$, where n denotes the sample size, ϕ is the persistence parameter, and $\varepsilon_t \sim \text{iid}\,(0, \sigma_\varepsilon^2)$. The SETAR(1,1) model is given by

$$x_t = \phi_1 x_{t-1} \mathbf{1}\{x_{t-d} > c\} + \phi_2 x_{t-1} \mathbf{1}\{x_{t-d} \leq c\} + \varepsilon_t,$$

where ϕ_1 and ϕ_2 are the persistence parameters of the first and second regime, respectively, $\mathbf{1}\{\cdot\}$ is the indicator function, d is the delay parameter, and c denotes the threshold. In this set-up an observation can be an outlier in the AR(1) process but a regular observation in the SETAR(1,1) model. This is the reason why linearity tests tend to spurious test decisions in the presence of outlying observations (cf. van Dijk et al., 1999).

To model outlier contaminated processes, the general replacement model of Martin and Yohai (1986) can be used. It divides the observable contaminated process y_t into an unobservable core process x_t and a contaminating process ζ_t,

$$y_t = x_t\,(1 - \delta_t) + \zeta_t\,\delta_t. \tag{3.1}$$

The AR(1) and SETAR(1,1) model form the unobservable core process x_t. The contaminating process ζ_t models AOs according to

$$\zeta_t = x_t + \zeta, \tag{3.2}$$

where ζ is the constant outlier magnitude. In order to model symmetric contaminations, the random variable δ_t takes the values 0 with the probability $1-\pi$ and ± 1 each with the probability $\pi/2$. The probability π is referred to as the outlier probability. Combining the definitions of Eq. (3.1) and (3.2), the contaminated process can be written as

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

Other specifications of the contaminating process ζ_t can be used to model other types of outliers, like innovative outliers, level shifts or temporary changes (cf. Galeano and Peña, 2013). However, the main focus in time series analysis is on AOs and innovative outliers as introduced by Fox (1972). According to van Dijk et al. (1999) and Ahmad and Donayre (2016) innovative outliers do not seriously deteriorate the performance of linearity tests. Therefore, this paper only considers AOs.

The effect of an AO is illustrated in Figure 3.1. The core process follows an AR(1) process with $\phi = 0.5$ and $\varepsilon_t \sim N(0, 1)$. An AO of magnitude $\zeta = 5$ occurs at observation $t = 50$. The contaminated process (black) and the core process (grey) only differ at $t = 50$ since the AO has no influence on the core process x_t. Thus, AOs only affect one single observation. If there were more outliers introduced in a linear process, they could be modeled as an additional regime, favoring the alternative hypothesis of a linearity test.

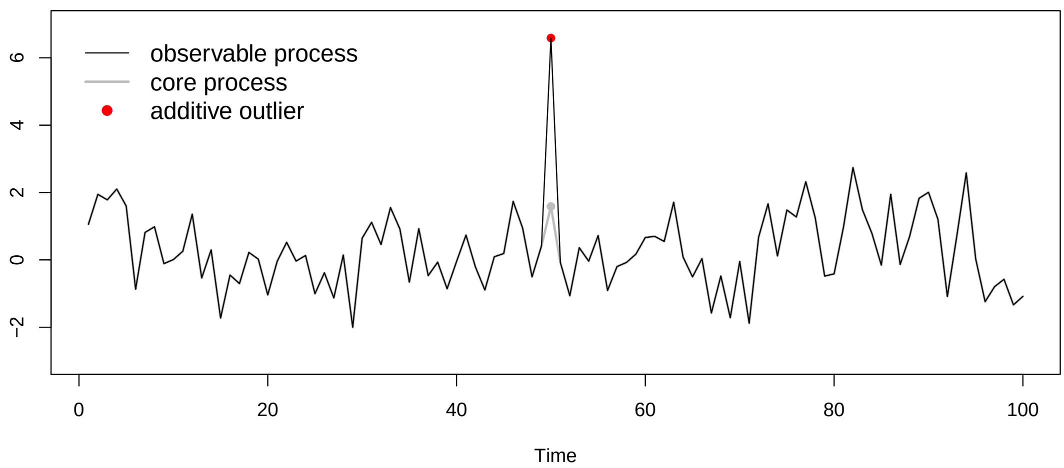


Figure 3.1: Effect of an Additive Outlier

3.3 Discriminating Linear and Nonlinear Models

There are two strands of procedures in order to discriminate between linear and nonlinear models, i.e. linearity tests and information criteria. There exists a variety of linearity tests designed to detect different types of nonlinearity. In this paper the focus is on the detection of SETAR nonlinearity. Therefore, the test against SETAR nonlinearity of Hansen (1999) is considered to compare the sensitivity of linearity tests and information criteria to AOs.

3.3.1 A Test against SETAR Nonlinearity

The test of Hansen (1999) is a F-type test to determine the number of regimes of a SETAR model. Since a SETAR model with one regime equals an AR process, the test can be used to detect SETAR nonlinearity. The focus of this paper is on the discrimination between linear and nonlinear models, not on lag order selection. Therefore, for simplicity the lag order is fixed at 1 and only AR and two-regime SETAR models are considered. These assumptions are not restrictive since a model order has to be determined before the application of the test, e.g. by using an information criterion. Therefore, only the first step of lag order determination is simplified in this set-up. The corresponding hypotheses of the test are given by

$$H_0 : y_t \sim \text{AR}(1) \qquad\qquad H_1 : y_t \sim \text{SETAR}(1,1)$$

or equivalently

$$H_0 : \phi_1 = \phi_2 \qquad\qquad H_1 : \phi_1 \neq \phi_2.$$

One of the major drawbacks of this test is the fact that both models have to be fully specified which increases the computational effort. Estimation is done by least-squares. In the case of the SETAR model, the autoregressive parameters are estimated conditionally on the threshold and on the delay. For both parameters a grid search is applied, for the threshold in the interval $[y_{0.15}, y_{0.85}]$ to ensure that at least 15% of the observations lie in each regime (cf. Hansen, 1997), and for the delay in the interval $[1, p]$ by convention (cf. Pitarakis, 2006). Since here p is fixed at 1, the delay is given by 1 as well. The parameter vector $\Big(c, d, \phi_1(c, d), \phi_2(c, d)\Big)$ that minimizes the residual sum of squares (RSS) of the SETAR model is selected. The test statistic is then given by

$$F = n \left(\frac{\text{RSS}_{\text{AR}} - \text{RSS}_{\text{SETAR}}}{\text{RSS}_{\text{SETAR}}} \right).$$

It depends on the estimated threshold $\hat{c}$ and on the estimated delay $\hat{d}$ through the RSS of the SETAR model. Since these parameters are only specified under the alternative, the test statistic does not converge asymptotically to a standard distribution. Therefore, critical values or p-values have to be simulated. In this paper the approach of Hansen (1999) is applied. Since critical values have to be determined under the null hypothesis, the estimated AR(1) model is the starting point of the simulation. So, residuals $\{e_t^*\}_{t=1}^n$ are sampled with replacement from the residuals of the AR(1) model. Under the initial condition $y_0^* = 0$ combined with the least squares estimate $\hat{\phi}$ from the original data, the simulated dependent variable y_t^* can be calculated recursively for $t = 1, \ldots, n$ according to

$$y_t^* = \hat{\phi} y_{t-1}^* + e_t^*.$$

For this simulated sample the test is conducted and the test statistic is saved. The procedure is repeated 2000 times and the empirical distribution function of the simulated test statistics is calculated. If the test statistic of the original sample exceeds the $(1-\alpha)$-quantile of the simulated test statistics, the null hypothesis will be rejected.

3.3.2 Information Criteria

Information criteria consist of a goodness of fit term and a penalty term to prevent overfitting. Different choices of the penalty term generate different information criteria with different characteristics. Generally, AIC (cf. Akaike, 1974), SIC (cf. Schwarz, 1978), and AICc (cf. Hurvich and Tsai, 1989) are applied. In this paper the focus will be on SIC and WIC (cf. Wu and Sepulveda, 1998) since AIC and AICc have a tendency to select the nonlinear model and hence are oversized, especially if the two regimes of the SETAR model are considered separately (cf. Gonzalo and Pitarakis, 2002; Pitarakis, 2006; Rinke and Sibbertsen, 2016),

$$\begin{aligned} \text{SIC} &= n\,\log(\hat{\sigma}^2) + p\,\log(n), \\ \text{WIC} &= n\,\log(\hat{\sigma}^2) + \frac{\Big(2n\,(p+1)/(n-p-2)\Big)^2 + \Big(p\,\log(n)\Big)^2}{2n\,(p+1)/(n-p-2) + p\,\log(n)}. \end{aligned}$$

The SIC tends to underfit in small samples. The WIC is a weighted version of AICc and SIC and is supposed to combine their advantages and thus perform well independent of the sample size.

For the SETAR model the two regimes are not considered separately, but the goodness of fit and the number of parameters are assessed for the whole model to calculate the values of the information criteria (cf. Rinke and Sibbertsen, 2016). Including the threshold as an additional parameter into the number of parameters of the SETAR model and applying the unbiased error term variance estimator

$$\hat{\sigma}^2 = \frac{\text{RSS}}{n-p-1}$$

prevents overfitting (cf. Kapetanios, 2001; Rinke and Sibbertsen, 2016; McQuarrie et al., 1997). The delay is not considered to be an additional parameter since it is fixed at 1 and does not have to be estimated.

The information criteria are calculated for both models, AR(1) and SETAR(1,1). Again, this implies a high computational effort, because both models have to be fitted to the data by (conditional) least-squares, similar to the estimation procedure of the linearity test. Given the parameter estimates, $\hat{\phi}$ for the AR(1) and $\hat{\phi}_1, \hat{\phi}_2, \hat{c}$, and $\hat{d}$ for the SETAR(1,1), the corresponding RSS, $\hat{\sigma}^2$, and the values of the information criteria can be calculated and a model can be selected. The linear model is selected if $\text{IC}_{\text{AR}} < \text{IC}_{\text{SETAR}}$.

3.4 Simulation Study

In the simulation study AR(1) and SETAR(1,1) models with different degrees of persistence $\phi, \phi_1, \phi_2 = \{\pm 0.75, \pm 0.50, \pm 0.25, 0.00\}$, different outlier magnitudes $\zeta = \{0, 1, 2, 3, 5\}$, and different sample sizes $n = \{100, 250, 500, 1000\}$ are considered. Throughout, the error terms ε_t are Gaussian white noise, the delay is fixed at 1, the threshold equals 0, and the outlier probability is $\pi = 0.05$ (cf. Ahmad and Donayre, 2016). Every series has a burn-in period of 200 observations to avoid a starting value bias. The initial values are set to zero. The simulation results are based on 1000 replications.

3.4.1 Size Properties

Let the size of information criterion be defined as the probability of selecting the nonlinear SETAR(1,1) model if the true DGP is a linear AR(1) process. The upper panel of Table 3.1 tabulates the size properties of the Hansen (1999) test for the typical significance levels $\alpha = \{1\%, 5\%, 10\%\}$ and of the SIC and the WIC for different degrees of persistence ϕ.

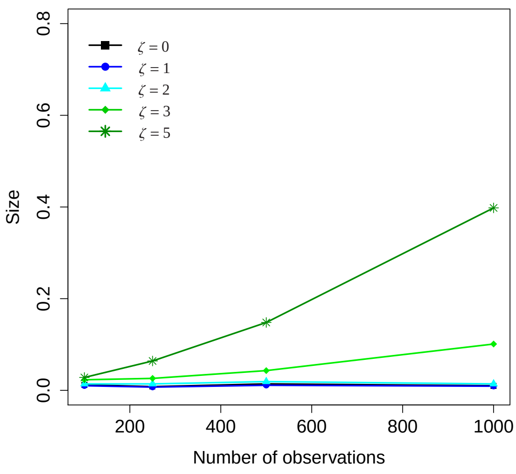
		$\phi = \mathbf{0.25}$					$\phi = \mathbf{0.75}$				
	n	1%	5%	10%	SIC	WIC	1%	5%	10%	SIC	WIC
$\zeta = \mathbf{0}$	**100**	0.012	0.056	0.111	0.054	0.114	0.011	0.053	0.101	0.048	0.102
	250	0.008	0.037	0.078	0.015	0.037	0.004	0.040	0.099	0.009	0.041
	500	0.014	0.057	0.094	0.015	0.039	0.009	0.042	0.102	0.010	0.023
	1000	0.010	0.044	0.089	0.004	0.014	0.015	0.060	0.102	0.006	0.024
$\zeta = \mathbf{3}$	**100**	0.023	0.071	0.125	0.069	0.130	0.065	0.185	0.282	0.172	0.276
	250	0.026	0.097	0.180	0.050	0.101	0.150	0.318	0.455	0.202	0.311
	500	0.043	0.149	0.252	0.045	0.096	0.292	0.511	0.636	0.290	0.430
	1000	0.101	0.307	0.450	0.060	0.144	0.554	0.781	0.867	0.454	0.621
$\zeta = \mathbf{5}$	**100**	0.028	0.092	0.162	0.088	0.160	0.241	0.540	0.679	0.518	0.675
	250	0.064	0.201	0.303	0.102	0.198	0.748	0.891	0.945	0.801	0.893
	500	0.148	0.389	0.520	0.147	0.273	0.982	0.999	0.999	0.982	0.997
	1000	0.398	0.659	0.787	0.304	0.475	1.000	1.000	1.000	1.000	1.000

Table 3.1: Size Properties of the Hansen (1999) Test and the Information Criteria. The DGP is an AR(1) process with $\phi = \{0.25, 0.75\}$ and $\zeta = \{0, 3, 5\}$.

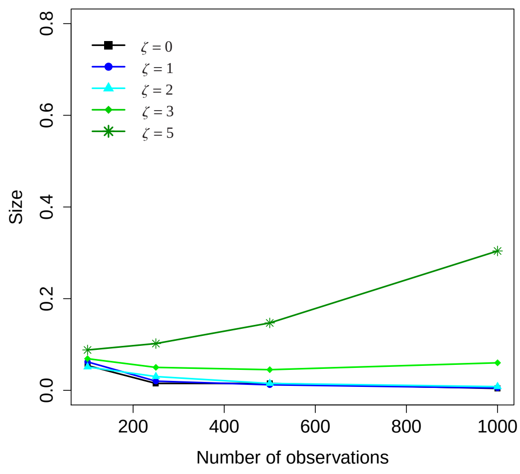
The linearity test has good size properties independent of the persistence and more importantly independent of the sample size due to the sample size dependent simulated critical values. In contrast, the size properties of the information criteria depend on the number of observations. So, in small samples the SIC and the WIC perform like the test at the 5% and the 10% level, respectively. For an increasing sample size the type I error of the information criteria decreases and converges towards zero since SIC and WIC are consistent (cf. Shibata, 1986; McQuarrie, 1999; Wu and Sepulveda, 1998). This can be seen in Figure 3.2 as well.

Figure 3.2 illustrates the effect of AOs with different outlier magnitudes $\zeta = \{0, 1, 2, 3, 5\}$ on the size of the linearity test and of the information criteria. In small samples $n = 100$ AOs only have a minor effect on the size of the linearity test and of the information criteria. This is probably due to the fact that there are only a few outliers (the expected number of outliers equals $n \cdot \pi$), which do not form a separate regime. The effect becomes more pronounced in larger samples. For small outlier magnitudes $\zeta = \{1, 2\}$ the size is not seriously deteriorated. However, for a larger outlier magnitude $\zeta = \{3, 5\}$ the Hansen (1999) test becomes seriously oversized. In contrast, the size of the information criteria decreases with the number of observations. Although the introduction of AOs increases the size again, the overall effect is less severe than for the linearity test. In large samples the SIC and the WIC outperform the test at the 1% and the 5% level, respectively.

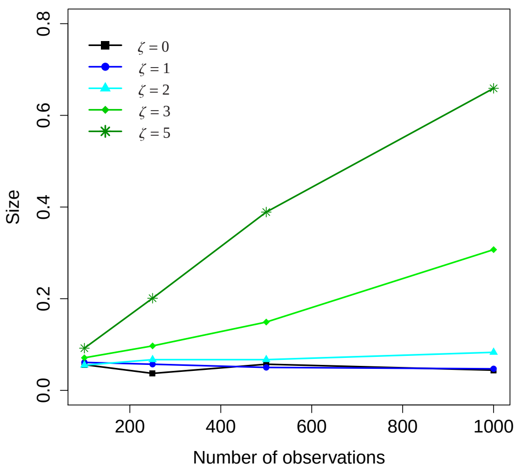
The effect of AOs on the size depends besides the outlier magnitude ζ and the sample size also on the degree of persistence. In the two lower panels of Table 3.1 the size of the Hansen (1999) test and of the information criteria is presented for different degrees of persistence ϕ and different outlier magnitudes ζ. The higher the degree of persistence, the higher are the size distortions. This coincides with the findings of Ahmad and Donayre (2016).



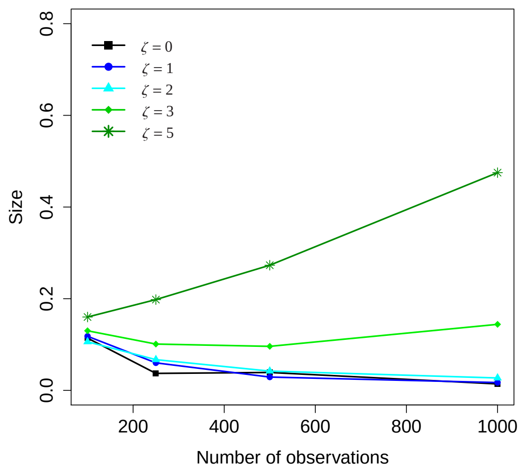
(a) $\alpha = 1\%$



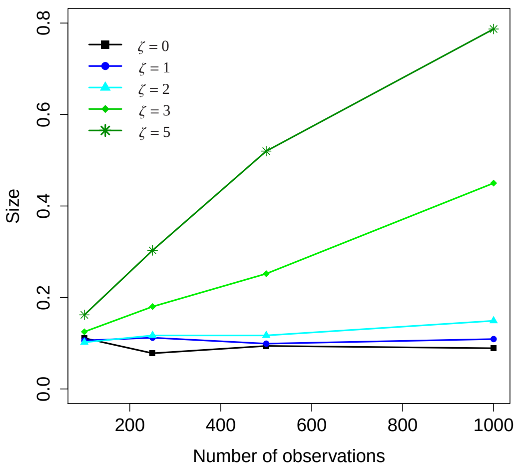
(b) SIC



(c) $\alpha = 5\%$



(d) WIC



(e) $\alpha = 10\%$

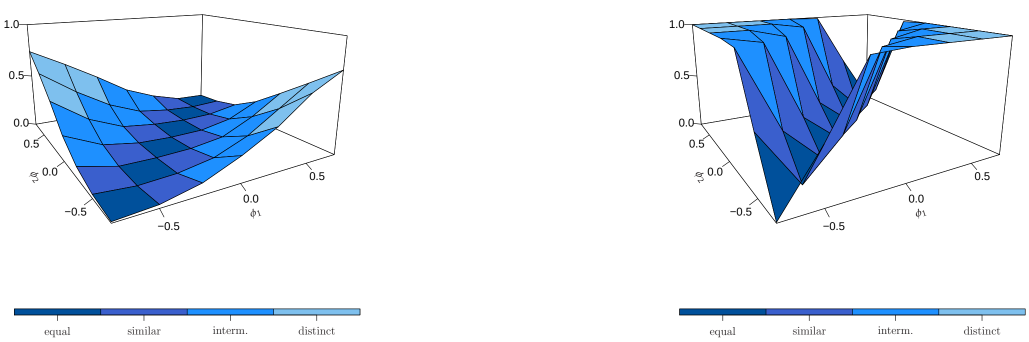
Figure 3.2: Size of the Hansen (1999) Test and the Information Criteria for Different Sample Sizes and Outlier Magnitudes ζ. The DGP is an AR(1) process with $\phi = 0.25$.

3.4.2 Power Properties

Let the power of an information criterion be defined as the probability of correctly selecting the SETAR(1,1) model. The power properties of the linearity test and of the information criteria depend on the sample size and on the difference between the regimes of the SETAR(1,1) model, i.e. the difference between the persistence parameters ϕ_1 and ϕ_2. To facilitate the interpretation of the power results, four different scenarios are distinguished depending on the absolute difference between ϕ_1 and ϕ_2.

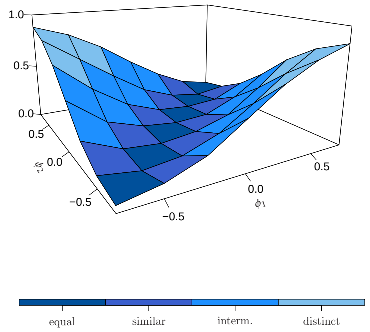
	Equal Regimes	Similar Regimes	Intermediate Regimes	Distinct Regimes
$\lvert\phi_1 - \phi_2\rvert$	0.00	0.25	$[0.50, 1.00]$	$[1.00, 1.50]$

The Power in Uncontaminated Series. The effect of the difference between the regimes and the sample size is illustrated in Figure 3.3. The more distinct the regimes, the higher is the power of the test and of the information criteria. For equal regimes the SETAR(1,1) model reduces to an AR(1) process. Thus, on the main diagonal the size is depicted. In small samples the power of the test and of the information criteria in similar regimes is relatively low. With an increasing number of observations the differentiation between linear and nonlinear models becomes easier and the power of the test and of the information criteria increases.

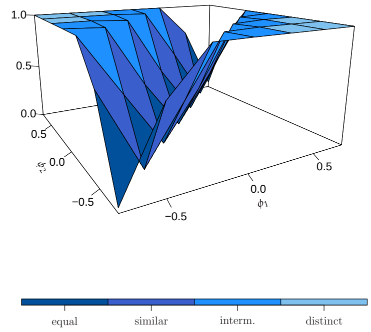


(a) $\alpha = 1\%$ and $n = 100$

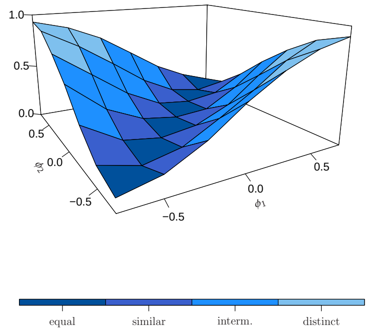
(b) $\alpha = 1\%$ and $n = 1000$



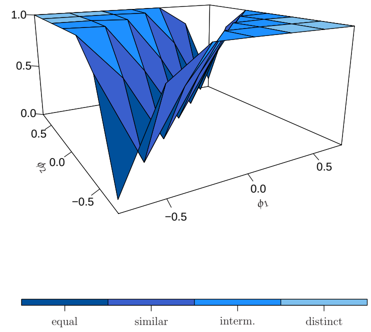
(c) $\alpha = 5\%$ and $n = 100$



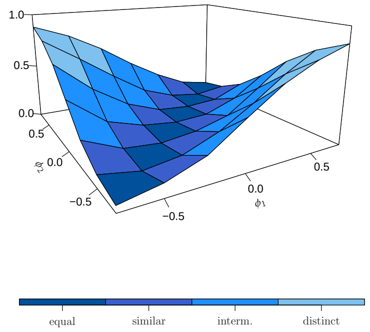
(d) $\alpha = 5\%$ and $n = 1000$



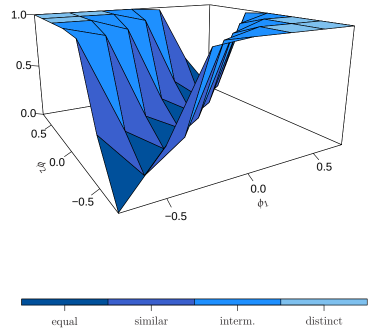
(e) $\alpha = 10\%$ and $n = 100$



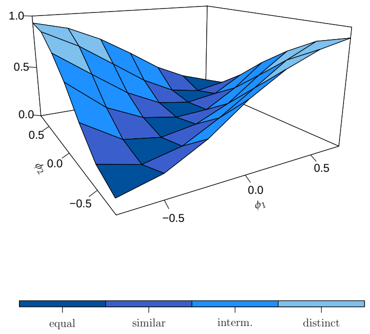
(f) $\alpha = 10\%$ and $n = 1000$



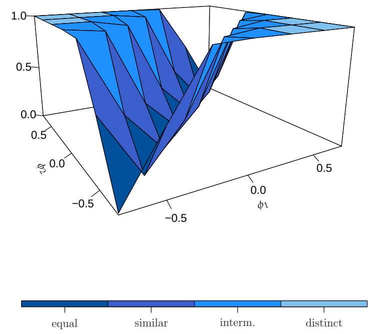
(g) SIC and $n = 100$



(h) SIC and $n = 1000$



(i) WIC and $n = 100$

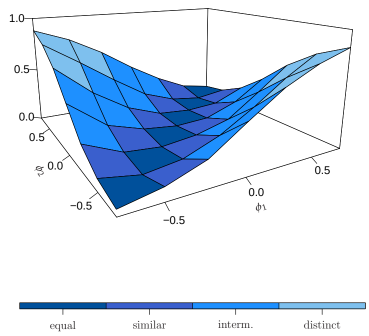


(j) WIC $n = 1000$

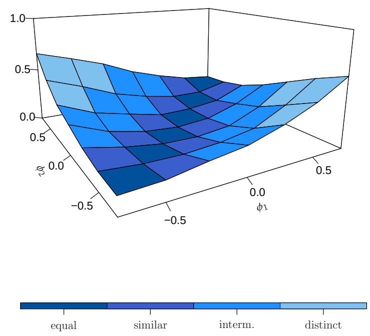
Figure 3.3: Power of the Hansen (1999) Test and the Information Criteria. The DGP is an uncontaminated SETAR(1,1) process.

The Power in Contaminated Series. The influence of the degree of persistence and of the outlier magnitude in contaminated series is pointed out in Figure 3.4 for the SIC. Again, the more distinct the regimes, the higher is the power of the SIC. However, the introduction of AOs changes the shape of the power plots. The size of the SIC increases, especially in highly persistent series. So, the power plot of the contaminated series with $\zeta = 5$ and $n = 100$ bends upwards at both ends of the diagonal. Therefore, the higher the absolute degree of persistence, the higher is the size distortion. In addition to increasing selection frequencies of the SETAR model in equal regimes, the power also increases in similar regimes with the outlier magnitude ζ and with the sample size. So, the differentiation between AR(1) and SETAR(1,1) in similar regimes becomes more reliable. Since the increase in power is at least partly due to the overall increase of the selection frequency of the nonlinear model, there exists a trade-off between power gains and size distortions.

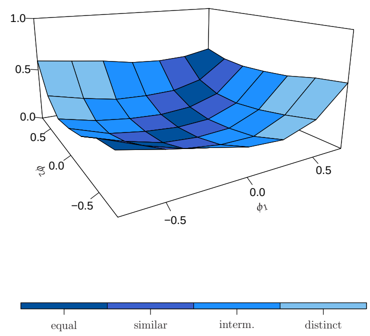
In small samples the power decreases in distinct regimes. With an increasing number of observations the power approaches 1, but then power losses occur in intermediate regimes.



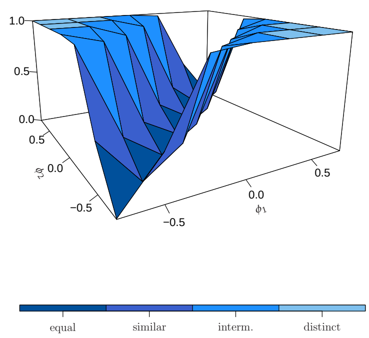
(a) $n = 100$ and $\zeta = 0$



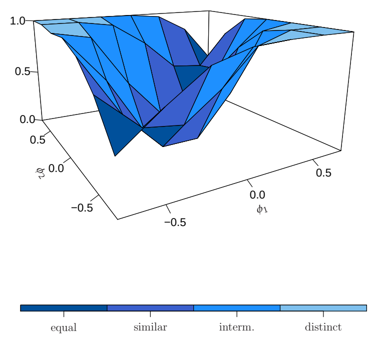
(b) $n = 100$ and $\zeta = 3$



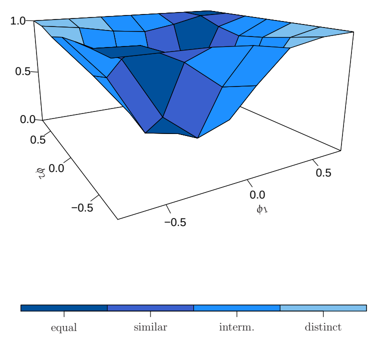
(c) $n = 100$ and $\zeta = 5$



(d) $n = 1000$ and $\zeta = 0$



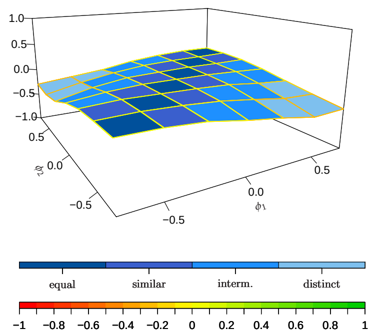
(e) $n = 1000$ and $\zeta = 3$



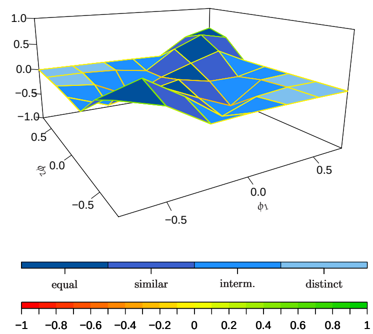
(f) $n = 1000$ and $\zeta = 5$

Figure 3.4: Power of the SIC. The DGP is a SETAR(1,1) process.

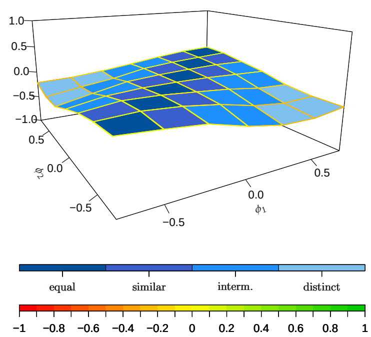
Since the power plots of the Hansen (1999) test and of the information criteria look very alike, the power plots of the contaminated series are only presented for the SIC. Instead, the change in power due to the introduction of AOs is visualized in Figures 3.5 and 3.6. The power for the outlier contaminated series with $\zeta = \{3, 5\}$ is compared to the uncontaminated series. Positive values indicate power gains, whereas negative values indicate power losses due to the contamination. The border color of the power plot facets and the corresponding red-green color bar help to quantify the change in power.



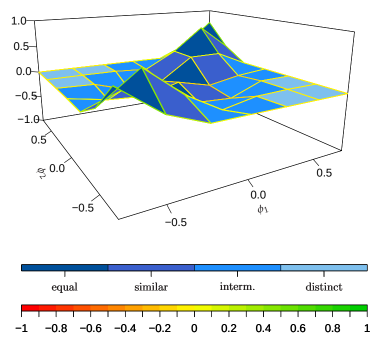
(a) $n = 100$ and $\alpha = 1\%$



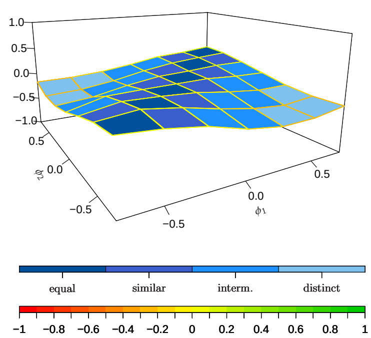
(b) $n = 1000$ and $\alpha = 1\%$



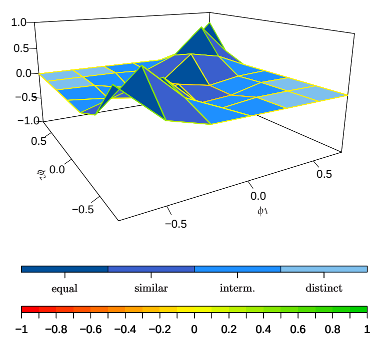
(c) $n = 100$ and $\alpha = 5\%$



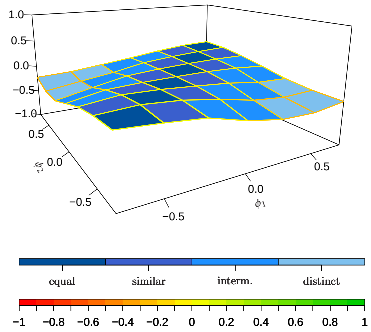
(d) $n = 1000$ and $\alpha = 5\%$



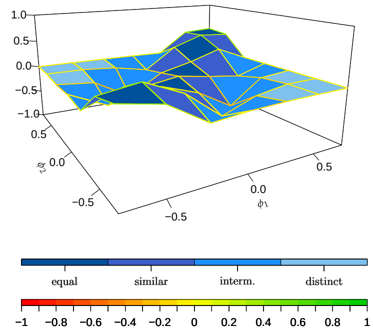
(e) $n = 100$ and $\alpha = 10\%$



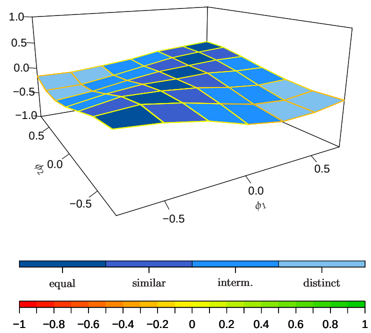
(f) $n = 1000$ and $\alpha = 10\%$



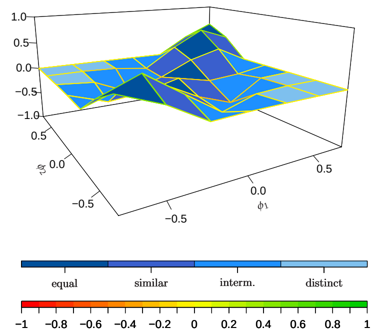
(g) $n = 100$ and SIC



(h) $n = 1000$ and SIC



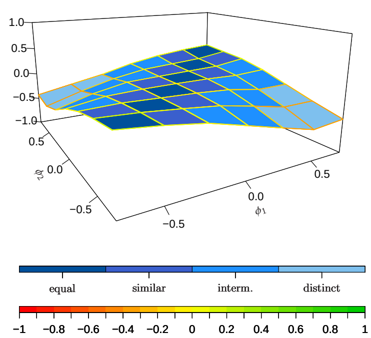
(i) $n = 100$ and WIC



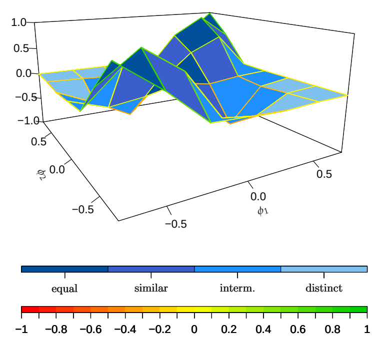
(j) $n = 1000$ and WIC

Figure 3.5: Change of Power for $\zeta = 3$ compared to $\zeta = 0$. The DGP is a SETAR(1,1) process.

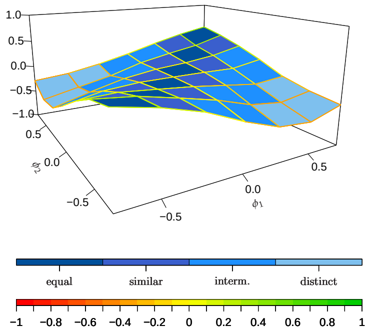
In small samples AOs lead to power gains in similar regimes, which also implies size distortions in equal regimes. In distinct regimes power losses occur. In large samples the main change is due to the size distortions. There occur no power losses in distinct regimes. However, there can be power losses in intermediate regimes, especially if the outlier magnitude increases (cf. Fig. 3.6).



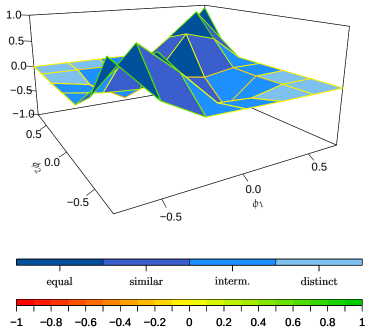
(a) $n = 100$ and $\alpha = 1\%$



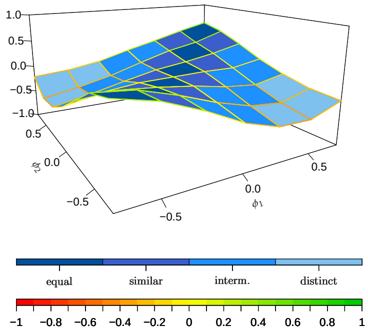
(b) $n = 1000$ and $\alpha = 1\%$



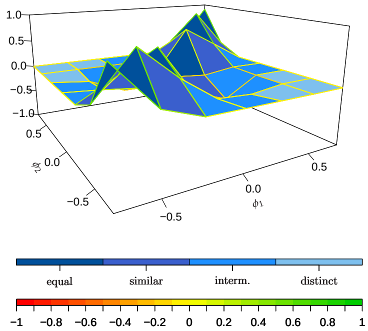
(c) $n = 100$ and $\alpha = 5\%$



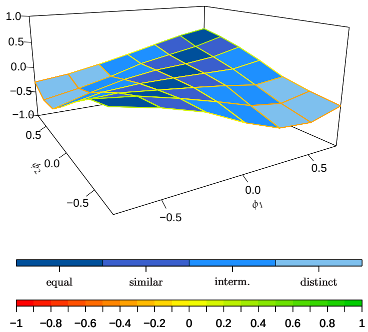
(d) $n = 1000$ and $\alpha = 5\%$



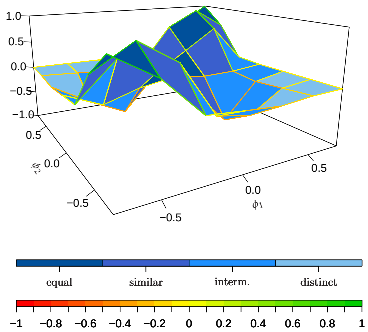
(e) $n = 100$ and $\alpha = 10\%$



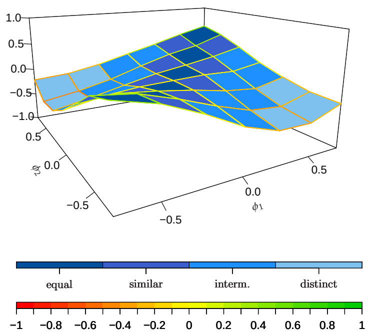
(f) $n = 1000$ and $\alpha = 10\%$



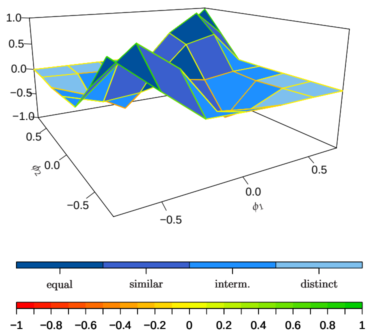
(g) $n = 100$ and SIC



(h) $n = 1000$ and SIC



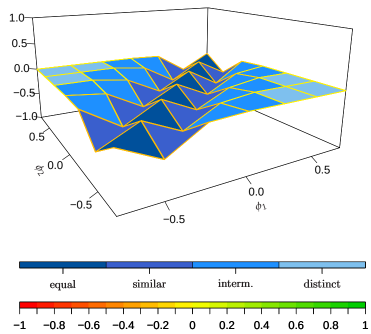
(i) $n = 100$ and WIC



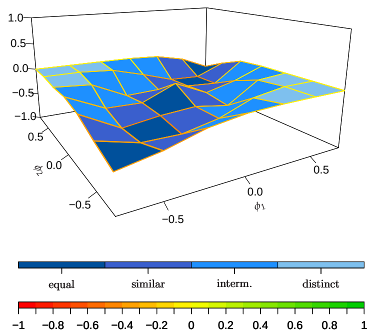
(j) $n = 1000$ and WIC

Figure 3.6: Change of Power for $\zeta = 5$ compared to $\zeta = 0$. The DGP is a SETAR(1,1) process.

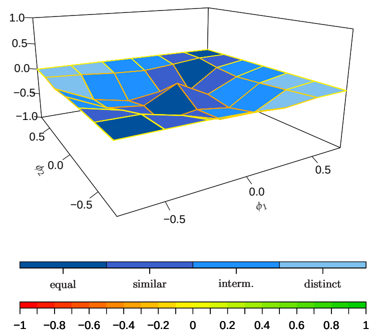
Comparison of the Hansen (1999) Test and the Information Criteria. The highest power is achieved by the Hansen (1999) test with $\alpha = 10\%$. In small samples the difference between the power of SIC and WIC compared to the test at the 10% level is negligible. Figure 3.7 depicts the power losses of the SIC and the WIC compared to the test at the 10% level for $n = 1000$.



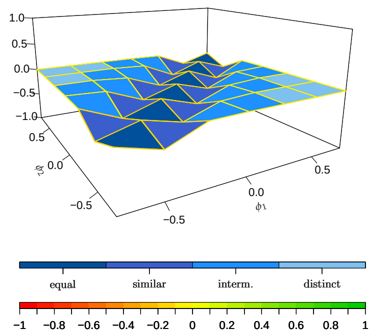
(a) SIC and $\zeta = 0$



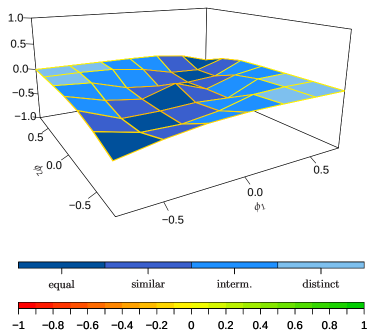
(b) SIC and $\zeta = 3$



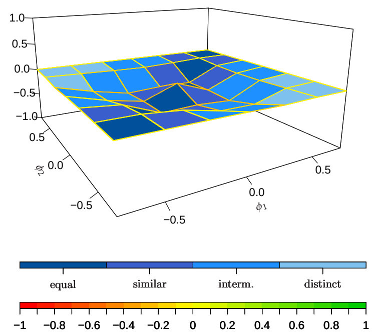
(c) SIC and $\zeta = 5$



(d) WIC and $\zeta = 0$



(e) WIC and $\zeta = 3$



(f) WIC and $\zeta = 5$

Figure 3.7: Comparison of the SIC and WIC with the Hansen (1999) Test at the 10% level. The DGP is a SETAR(1,1) with $n = 1000$.

In uncontaminated series the power loss of SIC and WIC occurs in similar regimes. For an increasing outlier magnitude the power loss increases in intermediate regimes. The power loss of the SIC is larger than that of the WIC. The "power loss" of the information criteria in equal regimes is actually the smaller size distortion. So, for the higher power of the test for large significance levels, severe size distortions have to be accepted. The power losses of the information criteria compared to the test approach zero with decreasing levels of significance.
In small samples SIC and WIC perform like the linearity test at the 5% and 10% significance level, respectively. In larger samples the information criteria perform approximately like the test at the 1% level.

3.5 Conclusion

In this paper the effects of additive outliers on the Hansen (1999) test against SETAR nonlinearity and on the information criteria SIC and WIC are investigated. AOs with a small outlier magnitude $\zeta = \{1, 2\}$ do not seriously deteriorate the performance of the test and the information criteria. For larger outlier magnitudes $\zeta = \{3, 5\}$ the size increases. Especially in persistent series and in larger samples test decisions become unreliable. Also the power can be negatively affected by large outliers. In small samples power losses occur in distinct regimes, in large samples in intermediate regimes. The effect of AOs on the power of the information criteria is similar to their effect on the Hansen (1999) test. In terms of size the results differ. In small samples of an uncontaminated series the size of SIC and WIC coincides with the size of the test at the 5% and 10% significance level, respectively. In contrast to the test, the size of the information criteria decreases with the number of observations and converges to zero. The size distortions in contaminated processes are less severe. In small samples the effect of AOs is not seriously deteriorating anyway, but in larger samples SIC and WIC are able to outperform the test at the 1% and 5% significance level, respectively. Due to their higher robustness against outlier contaminations in terms of size, the two information criteria are a valuable alternative to linearity tests.

Chapter 4

Changes in Persistence in Outlier Contaminated Time Series

Changes in Persistence in Outlier Contaminated Time Series

Co-authored with Tristan Hirsch

4.1 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. (1988b) 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 4.2 the model framework and the different outlier types are introduced. In Section 4.3 the tests for a change in persistence are explained. Section 4.4 introduces the outlier detection and removal methods. In Section 4.5 the simulation set-up and the simulation results are presented. Section 4.6 contains a real data example of the G7 inflation rates. Finally, Section 4.7 concludes.

4.2 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,

$$\Phi(L)\, x_t = \varepsilon_t, \qquad t = 1, \dots, T, \tag{4.1}$$

where T is the sample size, $\Phi(L) = 1 - \phi_1 L\,\mathbf{1}\{t \leq \lfloor \tau \cdot T \rfloor\} - \phi_2\, L\,\mathbf{1}\{t > \lfloor \tau \cdot T \rfloor\}$, 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),

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

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. (4.1).

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

$$\begin{aligned} &\zeta_t = x_t + \zeta, \\ \text{and} \quad &y_t = x_t + \zeta\delta_t, \end{aligned}$$

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

$$\begin{aligned} &\zeta_t = x_t + \zeta/\Phi(L) \\ \text{and} \quad &y_t = x_t + \Big(\zeta/\Phi(L)\Big)\, \delta_t. \end{aligned}$$

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.

4.3 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 4.5.

4.3.1 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 4.1 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 4.1: Simulated Critical Values of the Kim Test

4.3.2 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

$$R = \frac{K^f(\tau)}{K^r(\tau)} = \frac{\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} \hat{v}_{t,\tau}^2}{(T - \lfloor \tau T \rfloor)^{-2} \sum_{t=1}^{(T - \lfloor \tau T \rfloor)} \tilde{v}_{t,\tau}^2}, \qquad (4.2)$$

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 $I(0) \rightarrow I(1)$ 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. (4.2), 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 4.2 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 4.2: 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.

4.4 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 $\hat{\sigma}^2_\varepsilon$. 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.

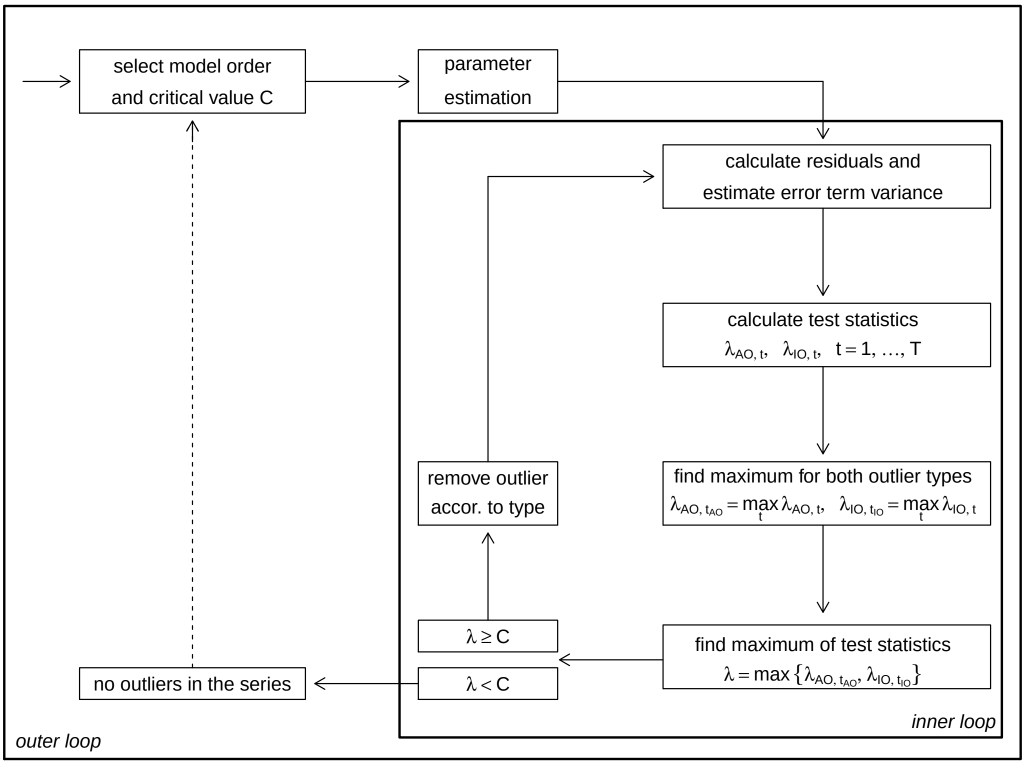


Figure 4.1: The Outlier Detection and Removal Method of Tsay (1988)

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.

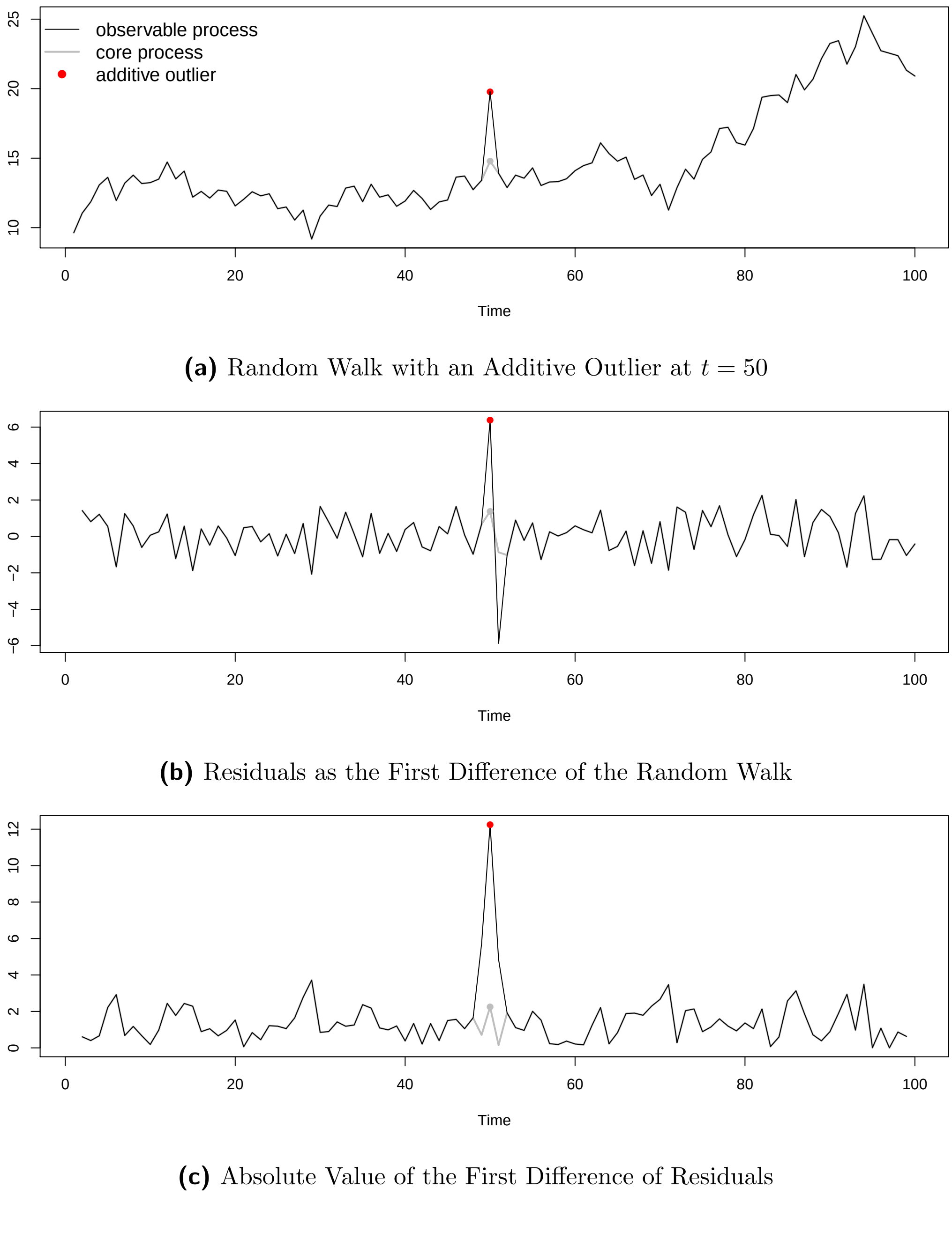
The described algorithm detects outliers sequentially, which is computationally easier and performs well if there exists only a single outlier in the series but can lead to biased estimates if there are multiple outliers (cf. Chen and Liu, 1993). Therefore, Chen and Liu (1993) propose a procedure consisting 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. Due to the fact that it is not known a priori when an AO occurs, the maximum of the absolute differences is determined. Let $t_{AO} = \arg\max_{2 \leq t \leq T-1} \left|e_{t+1} - e_t\right|$, 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

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

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 supplementary material.



(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

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.

4.5 Simulation Study

In our simulation study we consider the linear model given in Eq. (4.1) 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,

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

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.

4.5.1 Performance in Uncontaminated Series

Table 4.3 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. 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 4.4 tabulates the power results of the Kim test for $I(0) \rightarrow I(1)$ and of the Leybourne test

	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 4.3: Size Properties

for both $I(0) \rightarrow I(1)$ and $I(1) \rightarrow I(0)$. 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.

	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

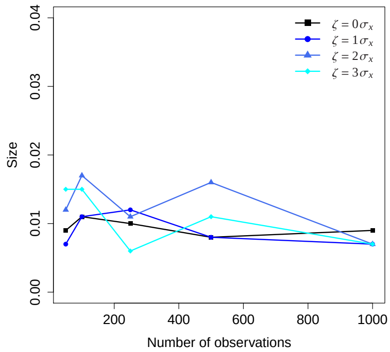
(c) Leybourne Test $I(0) \rightarrow I(1)$

Table 4.4: Power Properties

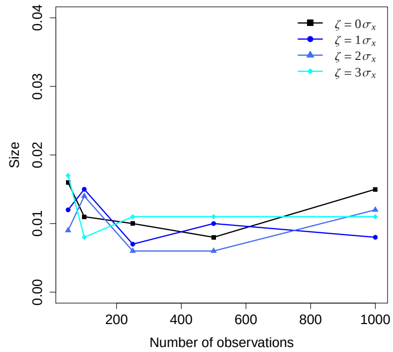
All presented results are valid for $\phi_1, \phi_2 = \{0, 1\}$. In general, the size of the Kim test increases if the degree of persistence increases and the power decreases with a decreasing change magnitude $|\phi_1 - \phi_2|$ (cf. Fig. 4.15 and 4.16). For the Leybourne test the size decreases to zero if the process becomes stationary. The power decreases if $|\phi_1 - \phi_2|$ decreases (cf. Fig. 4.17 and 4.18).

4.5.2 Performance in Contaminated Series

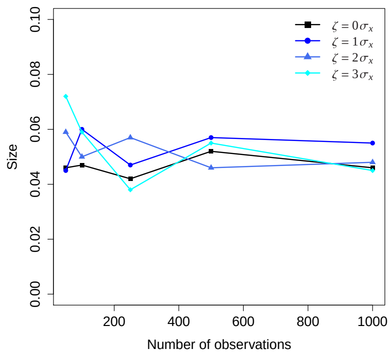
Figure 4.3 illustrates the effects of AOs and IOs on the size of the Kim test for different sample sizes, outlier magnitudes ζ, and significance levels. The results show that there is no difference between the effects of AOs and IOs on the size of the Kim test. This is due to the fact that the degree 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.



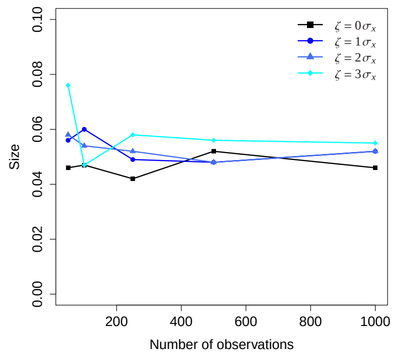
(a) AOs and $\alpha = 1\%$



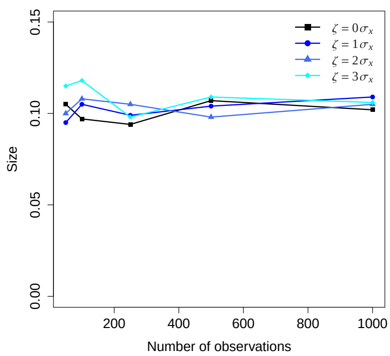
(b) IOs and $\alpha = 1\%$



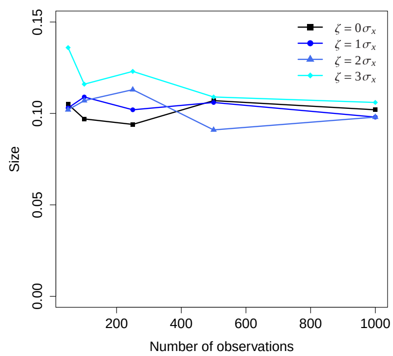
(c) AOs and $\alpha = 5\%$



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



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

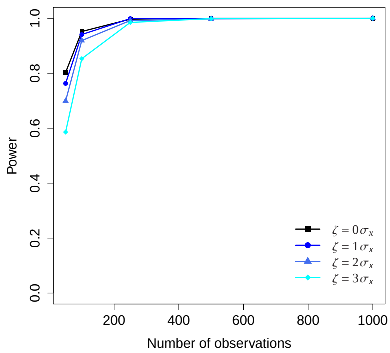


(f) IOs and $\alpha = 10\%$

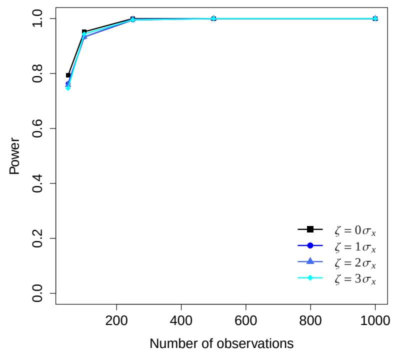
Figure 4.3: Size of the Kim Test ($I(0)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance

The higher the persistence of the simulated processes, the higher are the size distortions in small samples (cf. Fig. 4.15 and 4.16). However, the size is not deteriorated seriously, but holds the nominal significance level.

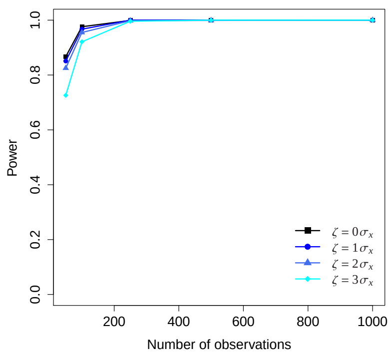
The power of the Kim test is not affected by AO contaminations if ζ is small. Only for large outlier magnitudes $\zeta = 3\sigma_x$ the power of the test decreases in small samples. The power of the test is not affected by IO contaminations (cf. Fig. 4.4).



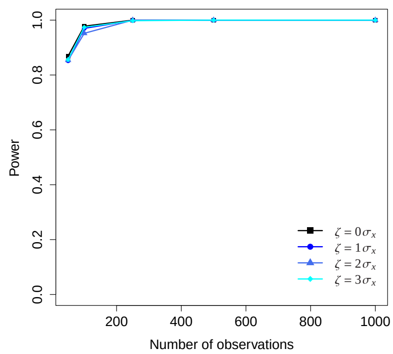
(a) AOs and $\alpha = 1\%$



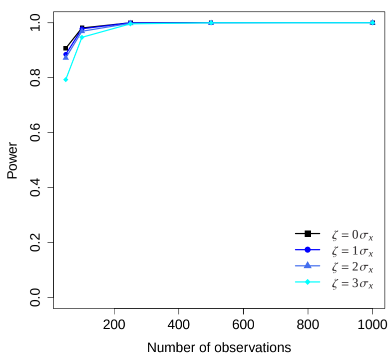
(b) IOs and $\alpha = 1\%$



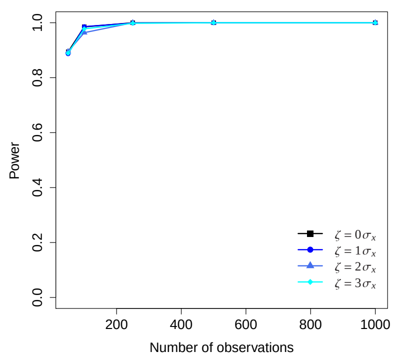
(c) AOs and $\alpha = 5\%$



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



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



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

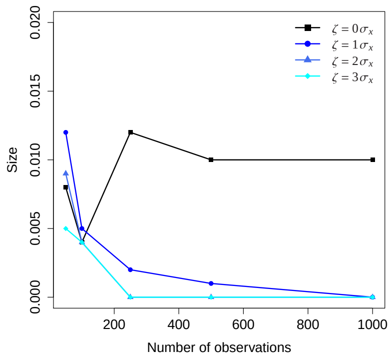
Figure 4.4: Power of the Kim Test ($I(0) \rightarrow I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance

Figure 4.5 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 4.5 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.

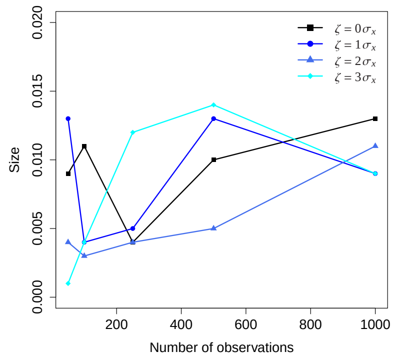
In terms of size the Leybourne test is more affected by outliers than the Kim test due to the higher degree of persistence under the null hypothesis. The effect of AOs is more serious than the effect of IOs.

Figures 4.6 and 4.7 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. 4.4). 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.

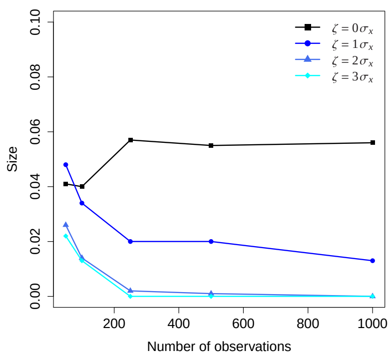
In contrast, IOs do not decrease the power, but lead to power gains since the stationary and the nonstationary part of the series markedly differ (cf. Fig. 4.8a).



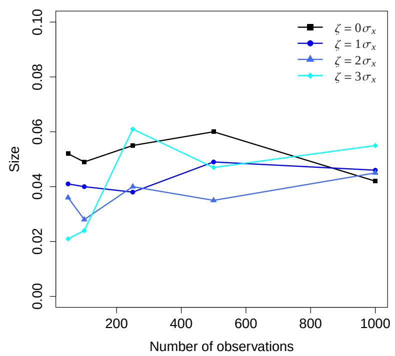
(a) AOs and $\alpha = 1\%$



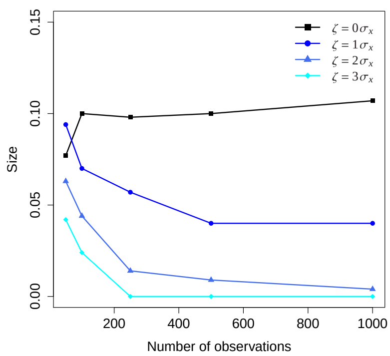
(b) IOs and $\alpha = 1\%$



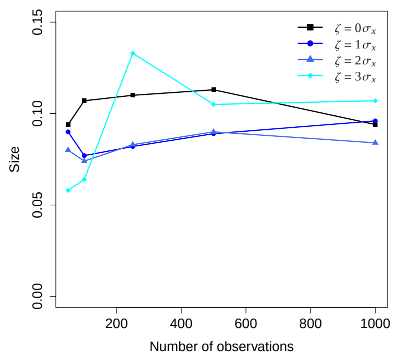
(c) AOs and $\alpha = 5\%$



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

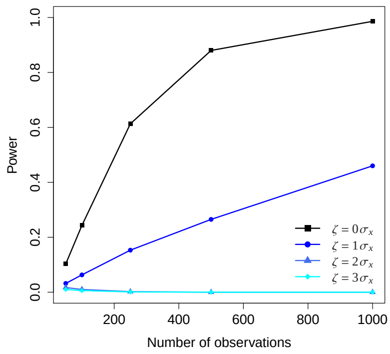


(e) AOs and $\alpha = 10\%$

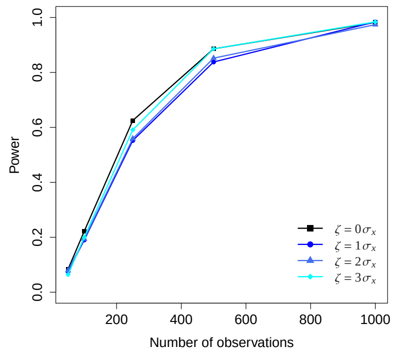


(f) IOs and $\alpha = 10\%$

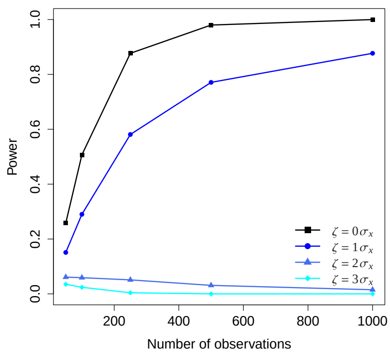
Figure 4.5: Size of the Leybourne Test ($I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance



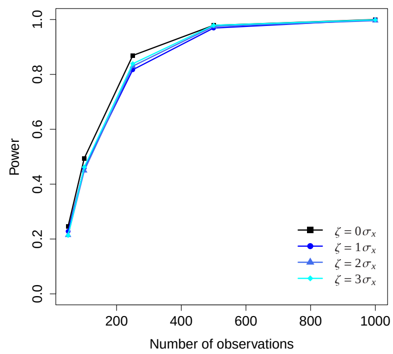
(a) AOs and $\alpha = 1\%$



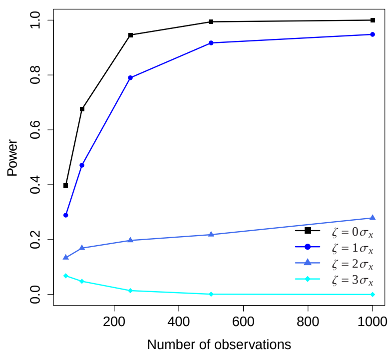
(b) IOs and $\alpha = 1\%$



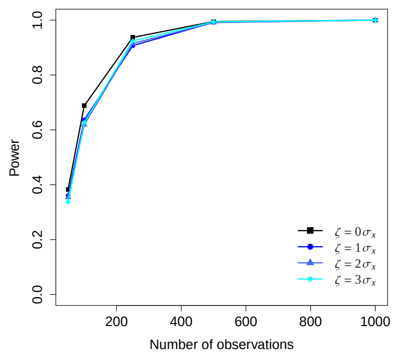
(c) AOs and $\alpha = 5\%$



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

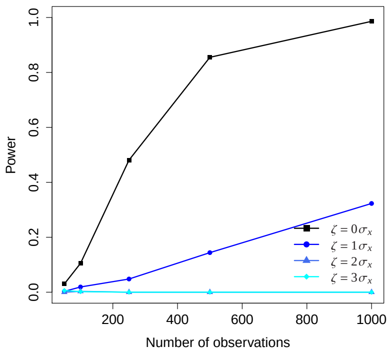


(e) AOs and $\alpha = 10\%$

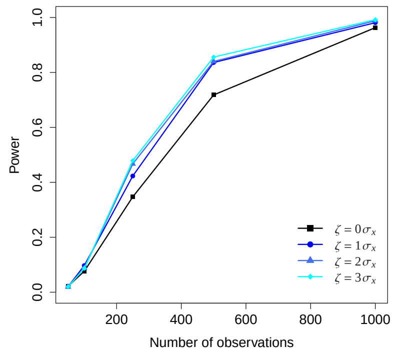


(f) IOs and $\alpha = 10\%$

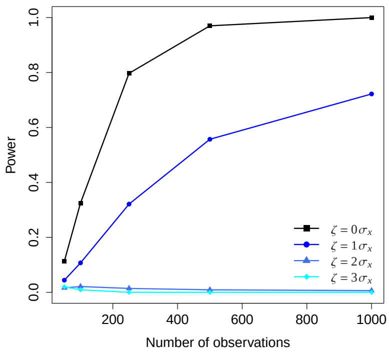
Figure 4.6: Power of the Leybourne Test ($I(0) \rightarrow I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance



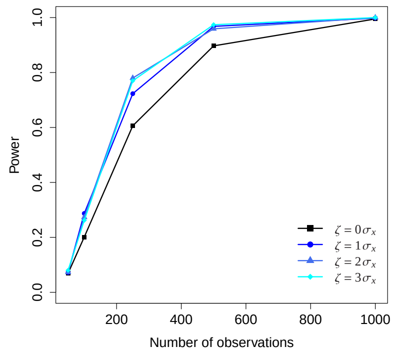
(a) AOs and $\alpha = 1\%$



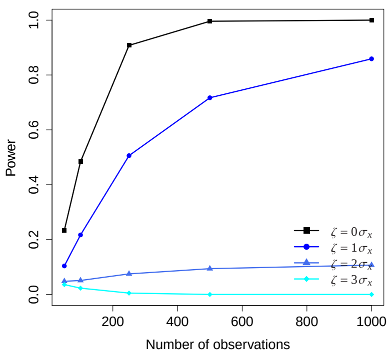
(b) IOs and $\alpha = 1\%$



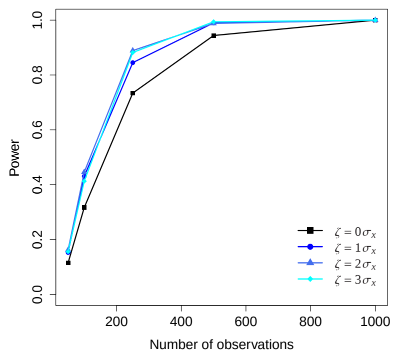
(c) AOs and $\alpha = 5\%$



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



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



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

Figure 4.7: 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

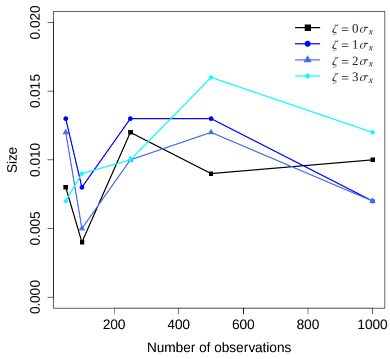
4.5.3 Performance in Adjusted Series

The results in Figures 4.3 and 4.4 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. 4.15 and 4.16).

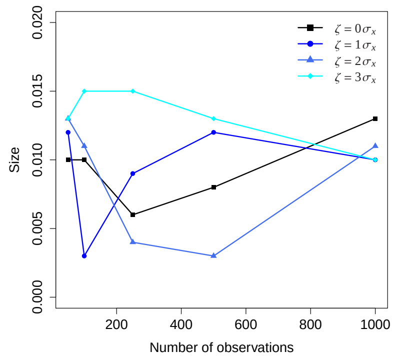
Figure 4.9 shows the size properties of the Leybourne test in the adjusted series. In all uncontaminated series the size is not affected by the adjustment procedure. Therefore, the algorithm does not spuriously detect outliers. Applying the modified algorithm of Shin et al. (1996) to AO contaminated 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 4.10 and 4.11 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. 4.8).

(a) Contaminated Series **(b)** Adjusted Series

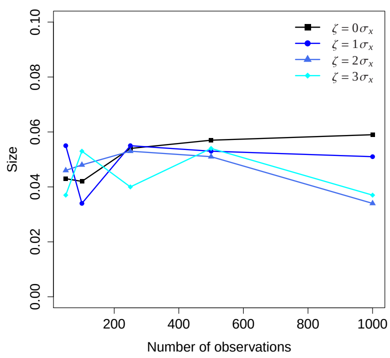
Figure 4.8: Influence of the Adjustment on an IO Contaminated Series with $I(0) \rightarrow I(1)$



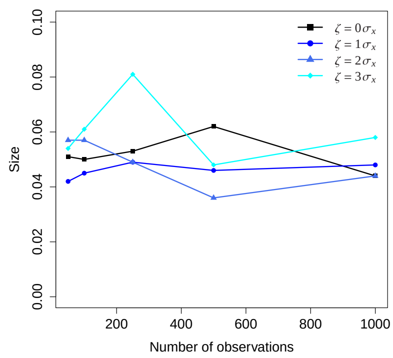
(a) AOs and $\alpha = 1\%$



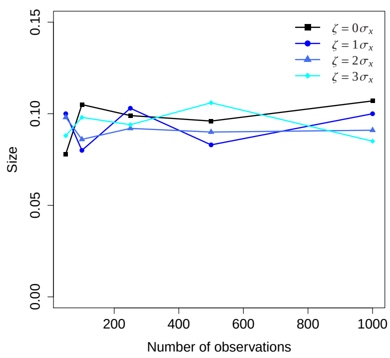
(b) IOs and $\alpha = 1\%$



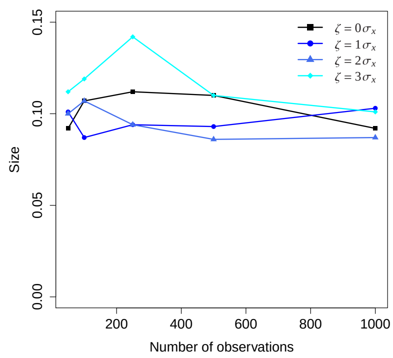
(c) AOs and $\alpha = 5\%$



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

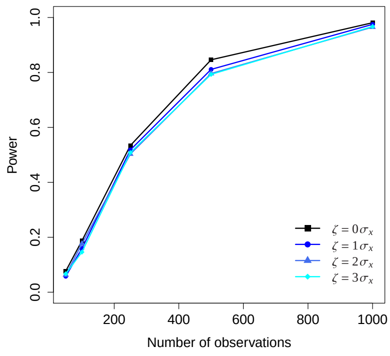


(e) AOs and $\alpha = 10\%$

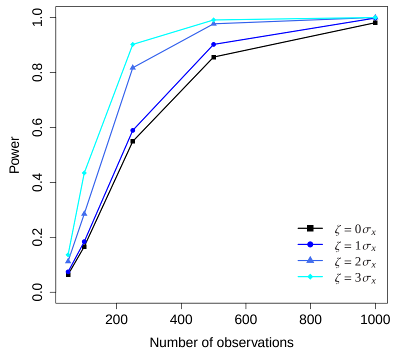


(f) IOs and $\alpha = 10\%$

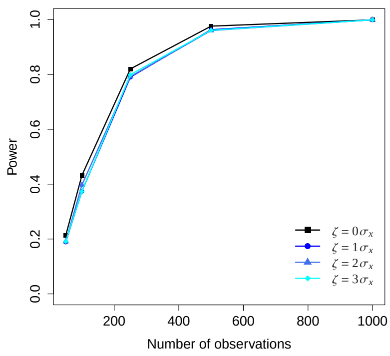
Figure 4.9: Size of the Leybourne Test ($I(1)$) for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance.



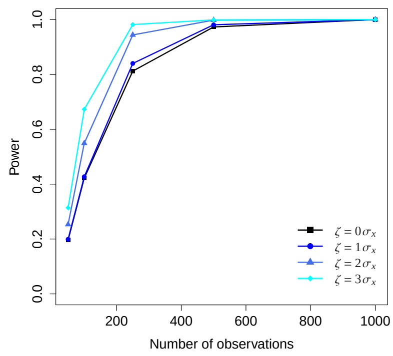
(a) AOs and $\alpha = 1\%$



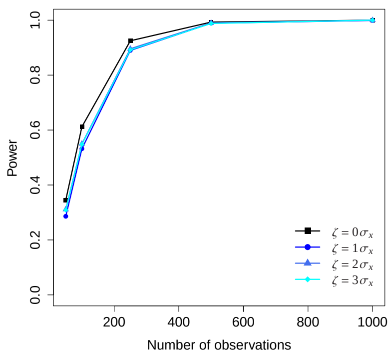
(b) IOs and $\alpha = 1\%$



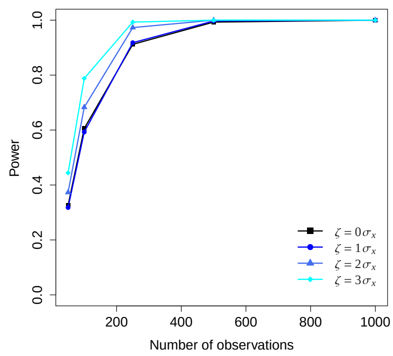
(c) AOs and $\alpha = 5\%$



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

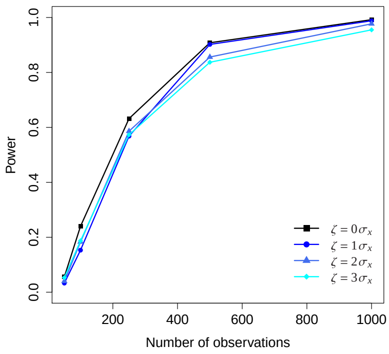


(e) AOs and $\alpha = 10\%$

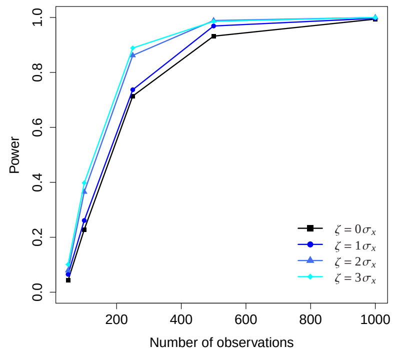


(f) IOs and $\alpha = 10\%$

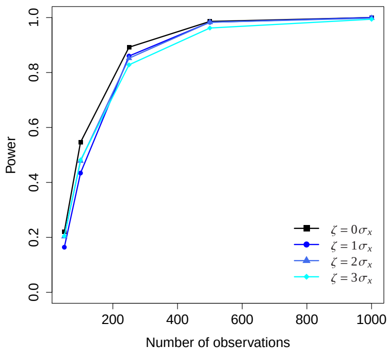
Figure 4.10: Power of the Leybourne Test $(I(0) \to I(1))$ for Additive and Innovative Outliers with Different Outlier Magnitudes ζ and Different Levels of Significance.



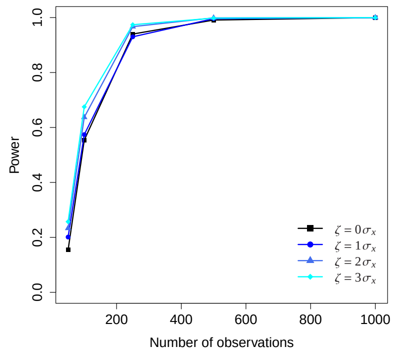
(a) AOs and $\alpha = 1\%$



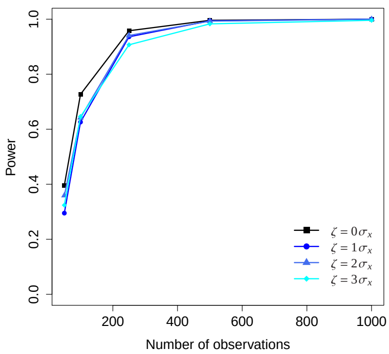
(b) IOs and $\alpha = 1\%$



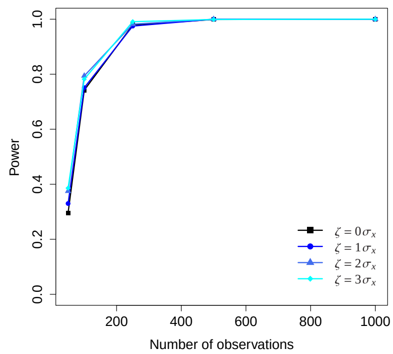
(c) AOs and $\alpha = 5\%$



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



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



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

Figure 4.11: 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.

4.6 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.

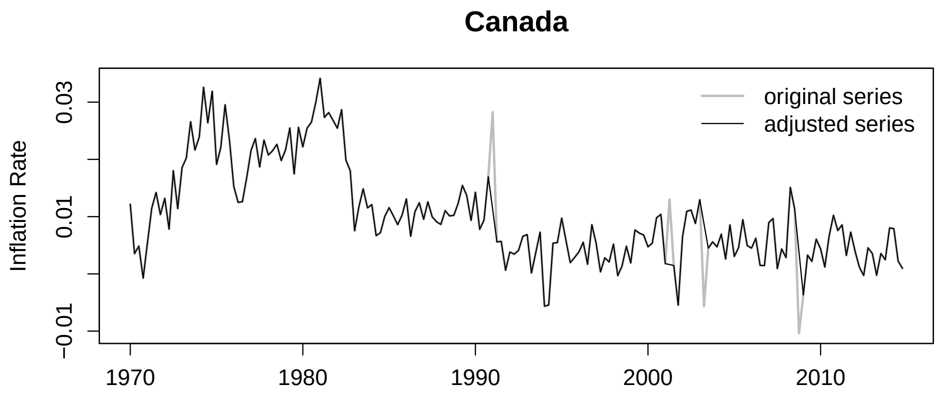
In Table 4.5 the critical values of both tests for a sample size of $T = 180$ are presented.

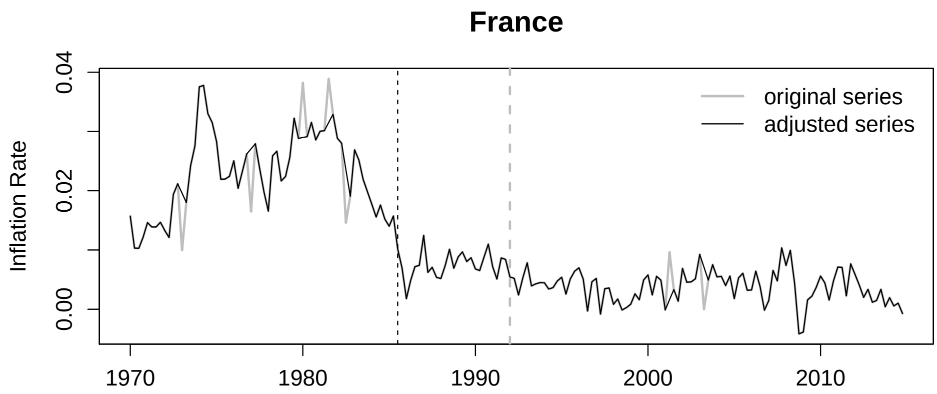
	Quantile					
	0.005	0.025	0.050	0.950	0.975	0.995
Kim	0.625	1.049	1.358	17.490	22.153	34.887
Leybourne	0.109	0.186	0.246	4.101	5.469	9.382

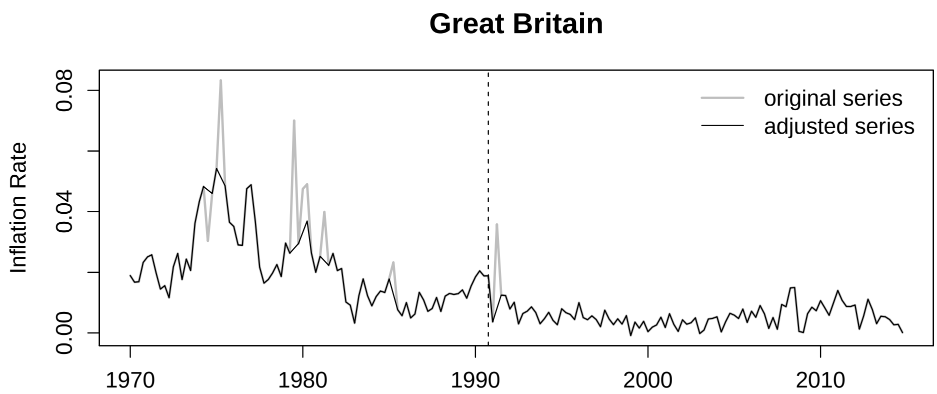
Table 4.5: Simulated Critical Values for $T = 180$

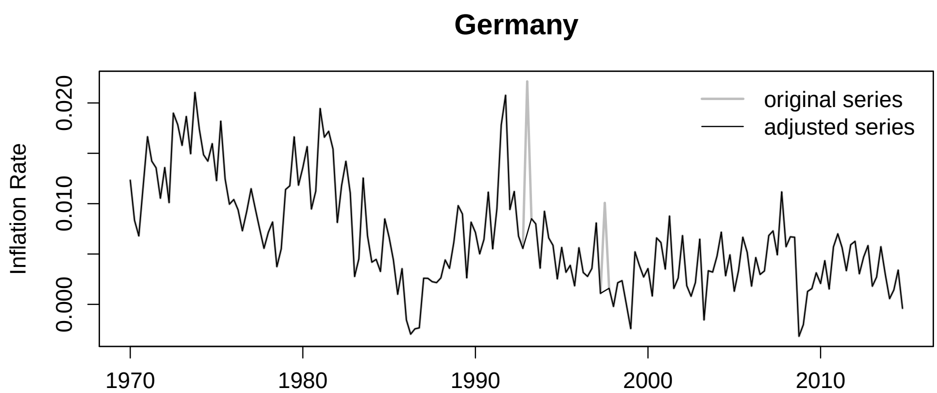
The test statistics of the Kim and Leybourne test applied to the G7 inflation rates for the original and the adjusted series are given in Table 4.6. Bold numbers indicate the rejection of the null hypothesis. In the original series the Kim test rejects the null hypothesis for Japan at the 10% significance 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.

In Figure 4.12 the original and the adjusted series of the G7 inflation rates are presented. The estimated change points are indicated by dashed lines.


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


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


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


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

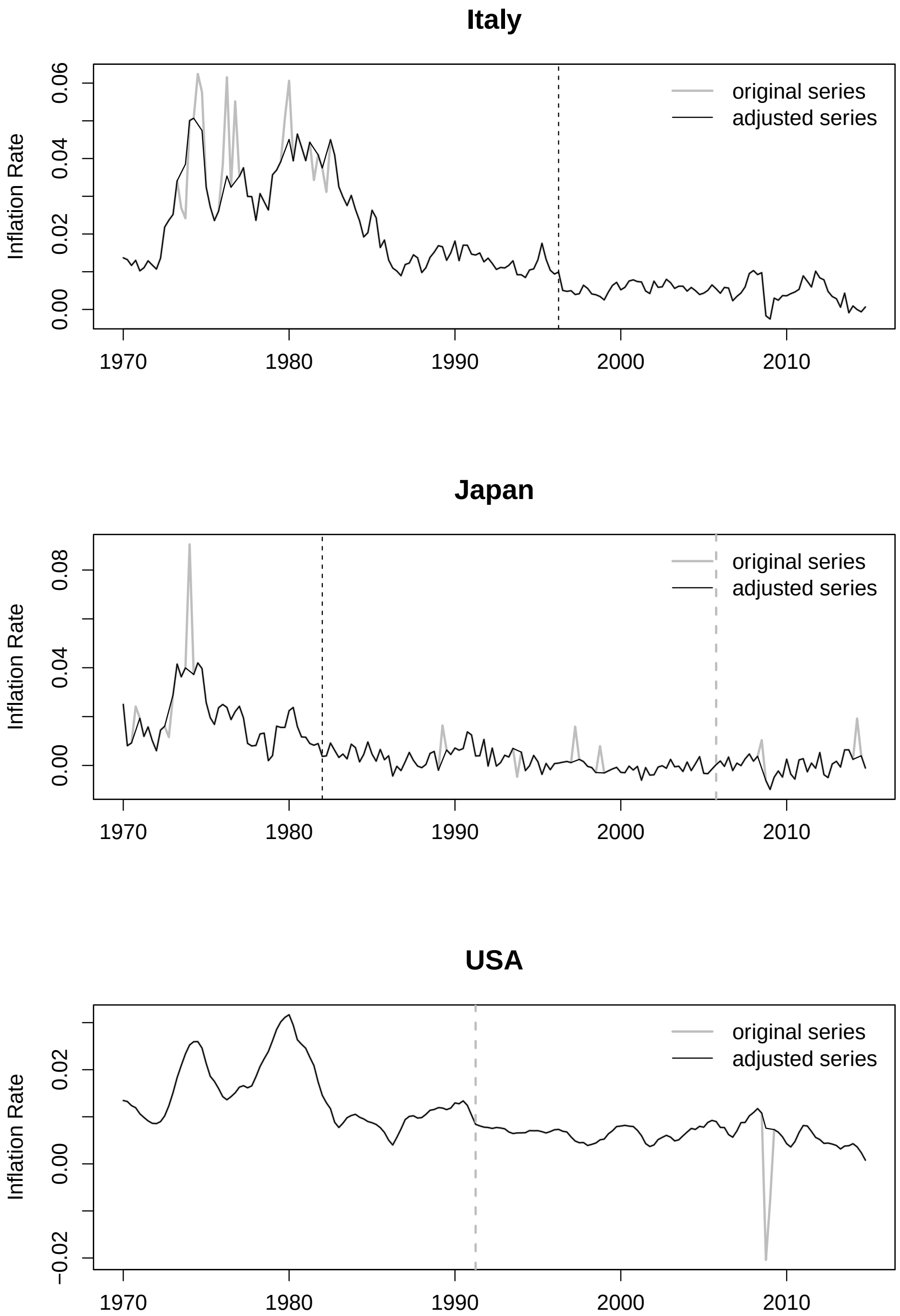


Figure 4.12: Inflation Rates of the G7 Countries

	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 4.6: Test Statistics of the Kim and Leybourne Test

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 subsamples. The p-values in Table 4.7 confirm the results of the tests for a 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 4.7: Subsample p-values of the ADF-Test

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 4.12. 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.

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.

4.7 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.

4.8 Supplementary Material

A 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,

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

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

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

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

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

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 4.13 and Table 4.8 illustrate the convergence of the test statistic λ^* to the standard folded normal distribution.

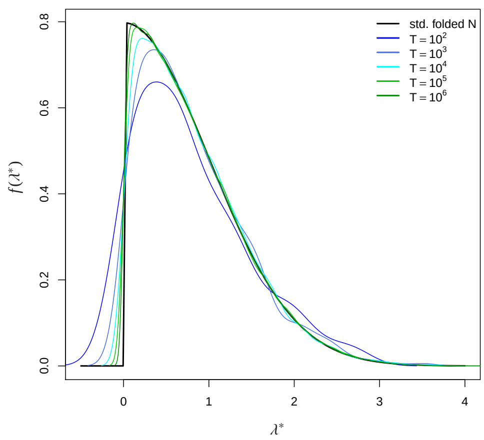


Figure 4.13: Estimated Density of λ^* and the Standard Folded Normal Distribution

	Quantile					
	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 4.8: 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,

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

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

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

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. 4.9), 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. 4.9).

	Quantile					
	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 4.9: 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 4.14.

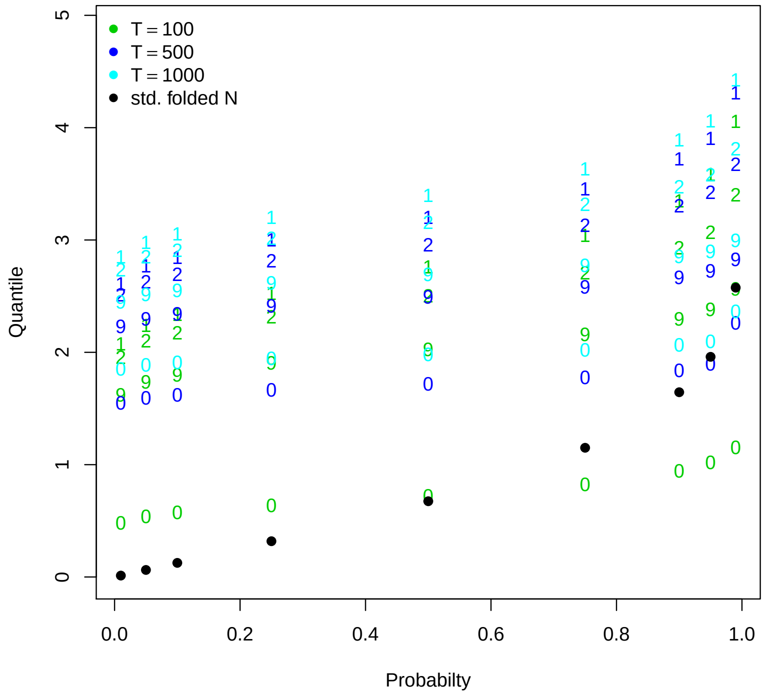
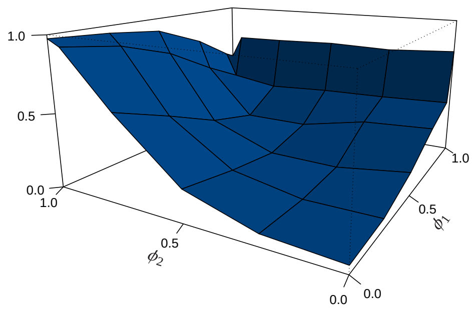


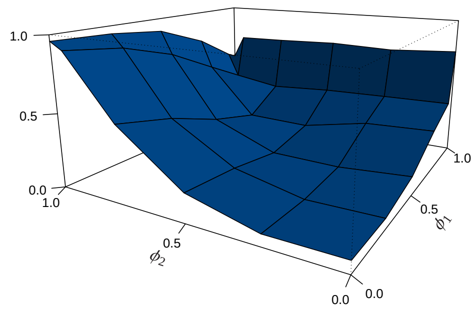
Figure 4.14: Estimated Quantiles of λ after Different Iterations

From Figure 4.14 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 4.14 the critical value of $C = 3$ is small enough for the test not to be conservative.

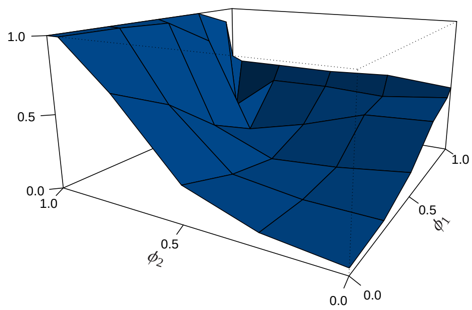
B Power Plots



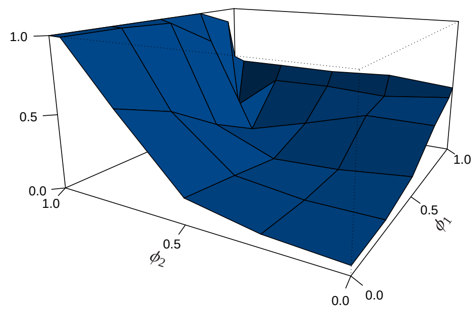
(a) Original Series, $T = 100$, $\zeta = 0$



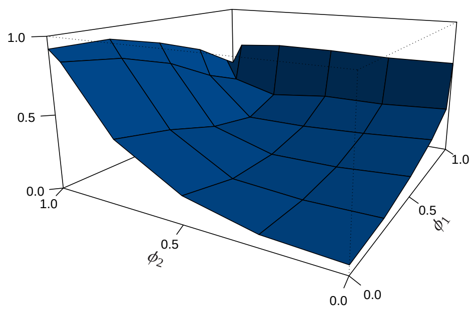
(b) Adjusted Series, $T = 100$, $\zeta = 0$



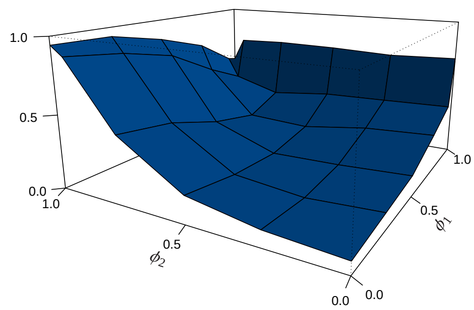
(c) Original Series, $T = 1000$, $\zeta = 0$



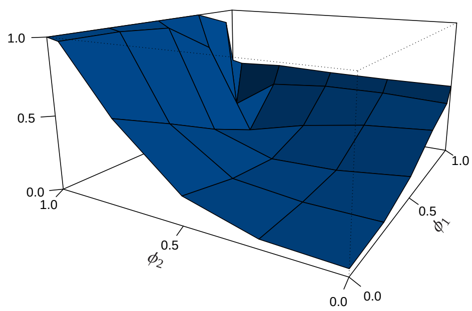
(d) Adjusted Series, $T = 1000$, $\zeta = 0$



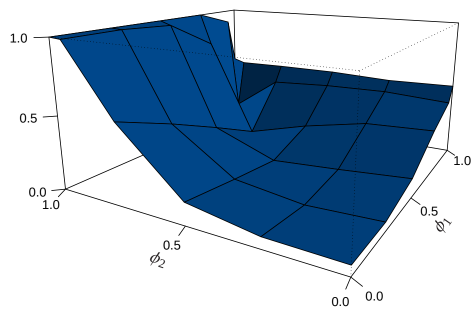
(e) Original Series, $T = 100$, $\zeta = 3$



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

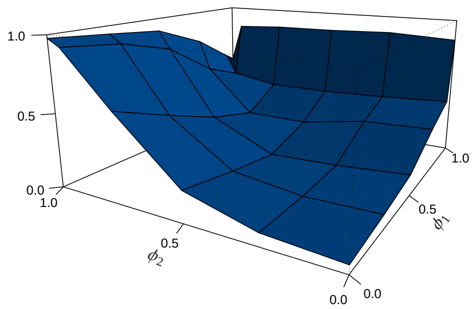


(g) Original Series, $T = 1000$, $\zeta = 3$

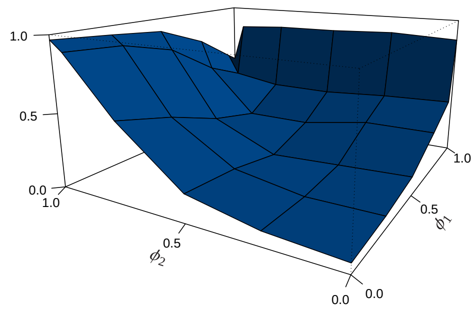


(h) Adjusted Series, $T = 1000$, $\zeta = 3$

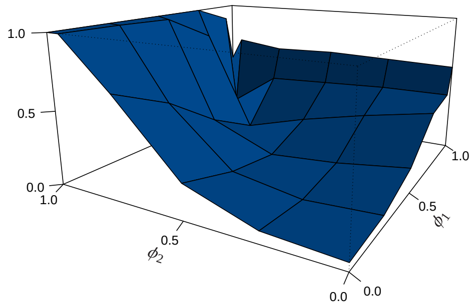
Figure 4.15: Power of the Kim Test for Additive Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.



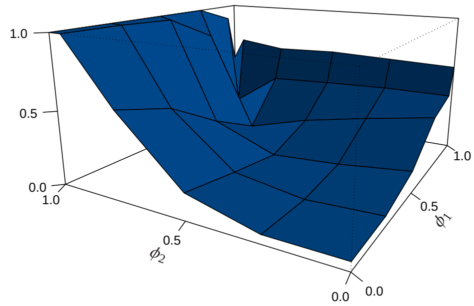
(a) Original Series, $T = 100$, $\zeta = 0$



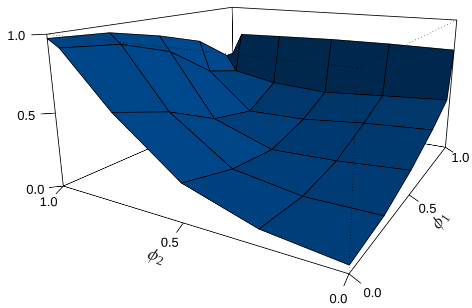
(b) Adjusted Series, $T = 100$, $\zeta = 0$



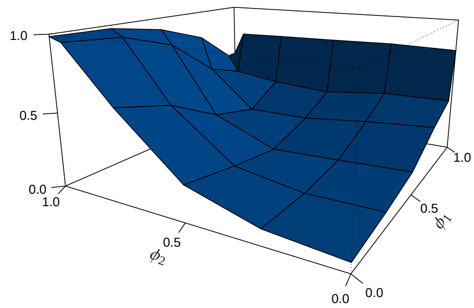
(c) Original Series, $T = 1000$, $\zeta = 0$



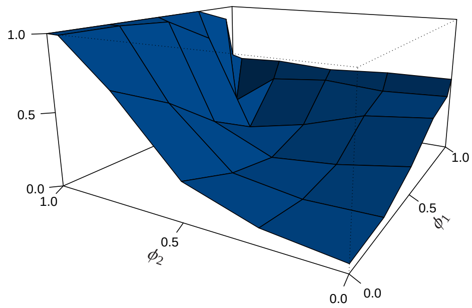
(d) Adjusted Series, $T = 1000$, $\zeta = 0$



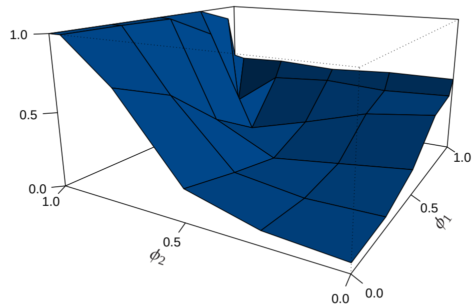
(e) Original Series, $T = 100$, $\zeta = 3$



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

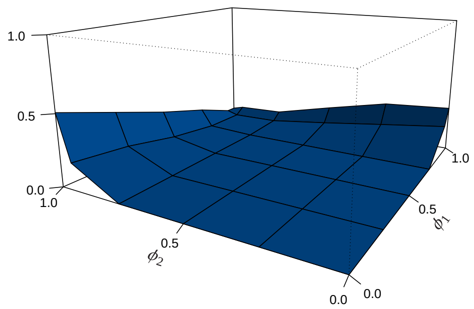


(g) Original Series, $T = 1000$, $\zeta = 3$

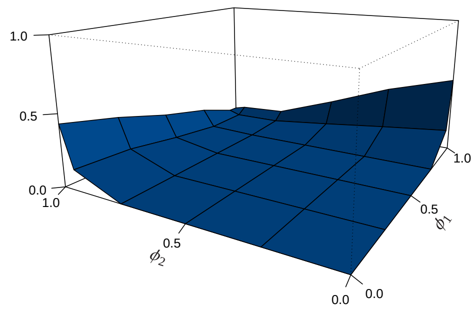


(h) Adjusted Series, $T = 1000$, $\zeta = 3$

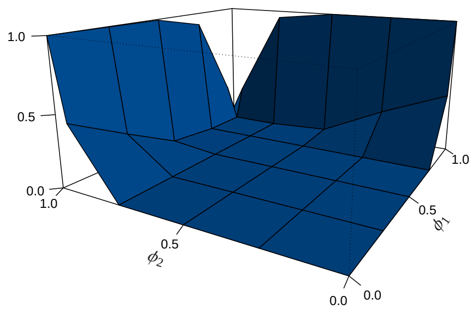
Figure 4.16: Power of the Kim Test for Innovative Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.



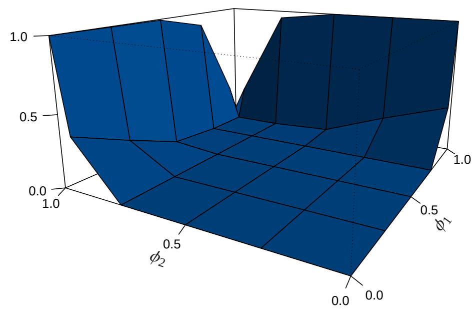
(a) Original Series, $T = 100$, $\zeta = 0$



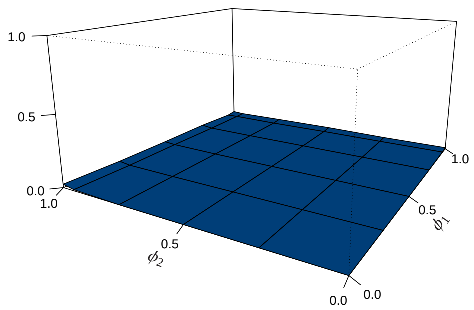
(b) Adjusted Series, $T = 100$, $\zeta = 0$



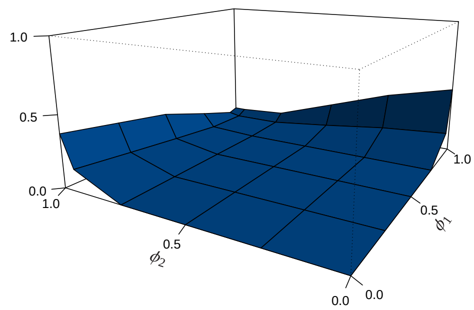
(c) Original Series, $T = 1000$, $\zeta = 0$



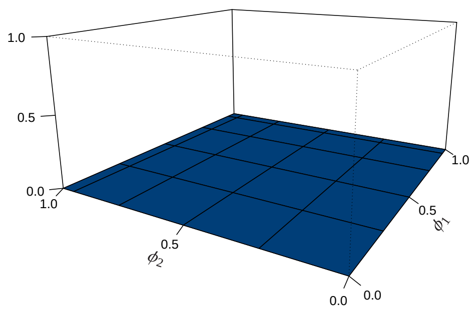
(d) Adjusted Series, $T = 1000$, $\zeta = 0$



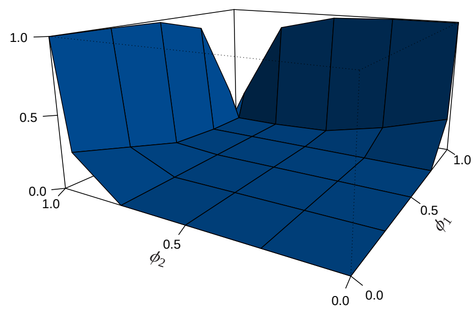
(e) Original Series, $T = 100$, $\zeta = 3$



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

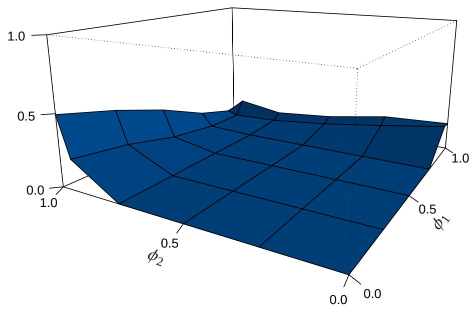


(g) Original Series, $T = 1000$, $\zeta = 3$

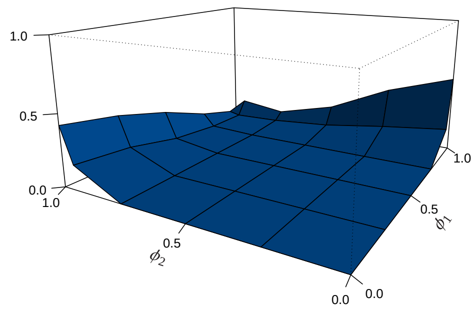


(h) Adjusted Series, $T = 1000$, $\zeta = 3$

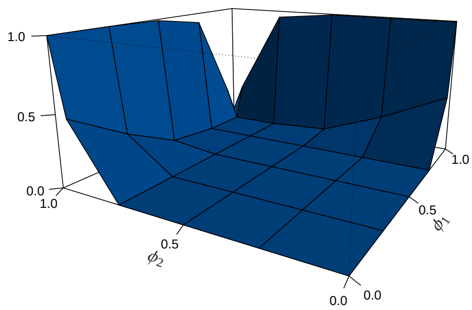
Figure 4.17: Power of the Leybourne Test for Additive Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.



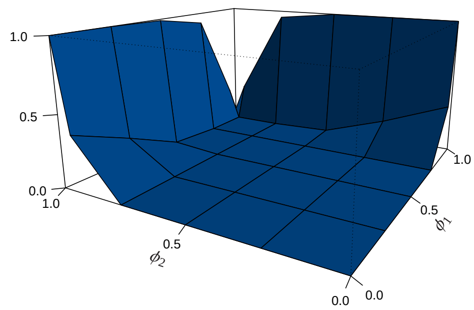
(a) Original Series, $T = 100$, $\zeta = 0$



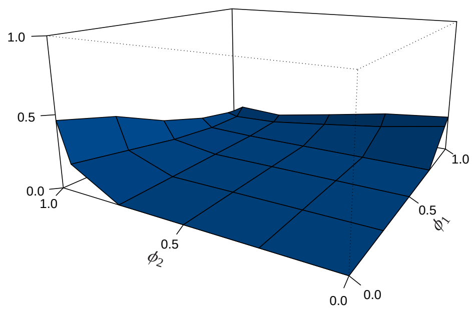
(b) Adjusted Series, $T = 100$, $\zeta = 0$



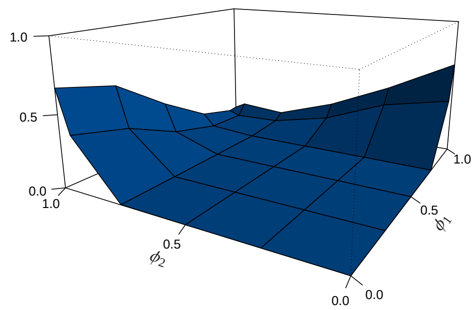
(c) Original Series, $T = 1000$, $\zeta = 0$



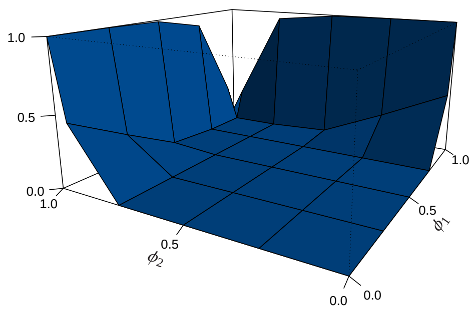
(d) Adjusted Series, $T = 1000$, $\zeta = 0$



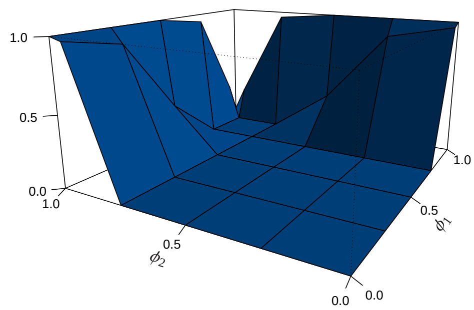
(e) Original Series, $T = 100$, $\zeta = 3$



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



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



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

Figure 4.18: Power of the Leybourne Test for Innovative Outliers, Different Degrees of Persistence, and $\alpha = 5\%$.

Bibliography

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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Box, G. E. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

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

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7(3):179–190.

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.

Clements, M. P. and Krolzig, H.-M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *The Econometrics Journal*, 1(1):47–75.

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(366):427–431.

Emiliano, P. C., Vivanco, M. J., and De Menezes, F. S. (2014). Information criteria: How do they behave in different models? *Computational Statistics & Data Analysis*, 69:141–153.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

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.

Gonzalo, J. and Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, 110(2):319–352.

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

Hamaker, E. (2009). Using information criteria to determine the number of regimes in threshold autoregressive models. *Journal of Mathematical Psychology*, 53(6):518–529.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton 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):1–14.

Hansen, B. E. (1999). Testing for linearity. *Journal of Economic Surveys*, 13(5):551–576.

Hughes, A. W. and King, M. L. (2003). Model selection using AIC in the presence of one-sided information. *Journal of Statistical Planning and Inference*, 115(2):397–411.

Hughes, A. W., King, M. L., and Kwek, K. T. (2004). Selecting the order of an ARCH model. *Economics Letters*, 83(2):269–275.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.

Kapetanios, G. (2001). Model selection in threshold models. *Journal of Time Series Analysis*, 22(6):733–754.

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.

Li, W. (1988). The Akaike information criterion in threshold modelling: Some empirical evidences. In *Nonlinear Time Series and Signal Processing*, volume 106 of *Lecture Notes in Control and Information Sciences*, pages 88–96. Springer Berlin Heidelberg.

Liu, J., Wu, S., and Zidek, J. V. (1997). On segmented multivariate regression. *Statistica Sinica*, 7(2):497–525.

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

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988b). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics*, 15(3):161–175.

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

McQuarrie, A., Shumway, R., and Tsai, C.-L. (1997). The model selection criterion AICu. *Statistics & Probability letters*, 34(3):285–292.

McQuarrie, A. D. (1999). A small-sample correction for the Schwarz SIC model selection criterion. *Statistics & Probability letters*, 44(1):79–86.

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

Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6):1361–1401.

Perron, P. (2006). Dealing with structural breaks. In *Palgrave Handbook of Econometrics - Econometric Theory*, chapter 8, pages 278 – 352. Palgrave Macmillan.

Pitarakis, J.-Y. (2006). Model selection uncertainty and detection of threshold effects. *Studies in Nonlinear Dynamics & Econometrics*, 10(1).

Psaradakis, Z., Sola, M., Spagnolo, F., and Spagnolo, N. (2009). Selecting nonlinear time series models using information criteria. *Journal of Time Series Analysis*, 30(4):369–394.

Rinke, S. and Sibbertsen, P. (2016). Information criteria for nonlinear time series models. *Studies in Nonlinear Dynamics & Econometrics*, 20(3):325–341.

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.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shibata, R. (1986). Consistency of model selection and parameter estimation. *Journal of Applied Probability*, 23:127–141.

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.

Smith, A., Naik, P. A., and Tsai, C.-L. (2006). Markov-switching model selection using Kullback–Leibler divergence. *Journal of Econometrics*, 134(2):553–577.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218.

Tong, H. (1983). *Threshold models in non-linear time series analysis. Lecture Notes in Statistics, No. 21.* Springer-Verlag.

Tong, H. (1990). *Non-linear time series: A dynamical system approach.* Oxford University Press.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(3):245–292.

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.

Wen, M.-J. and Tu, Y.-H. (2001). Modified WIC for order selection in autoregressive model. Technical Report No. 40, National Cheng-Kung University, Institute of Statistics.

Wong, C. S. and Li, W. K. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models. *Journal of Time Series Analysis*, 19(1):113–124.

Wu, T.-J. and Sepulveda, A. (1998). The weighted average information criterion for order selection in time series and regression models. *Statistics & Probability Letters*, 39(1):1–10.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 226(636-646):267–298.