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

Saskia Rinke[1]

Leibniz University Hannover

Abstract
In this paper the performance of information criteria and a test against SETAR nonlinearity for outlier contaminated time series are investigated. Additive outliers can seriously influence the properties of the underlying time series and hence of linearity tests, resulting in spurious test decisions of nonlinearity. Using simulation studies, the performance of the information criteria SIC and WIC as an alternative to linearity tests are assessed in time series with different degrees of persistence and different outlier magnitudes. For uncontaminated series and a small sample size the performance of SIC and WIC is similar to the performance of the linearity test at the 5% and 10% significance level, respectively. For an increasing number of observations the size of SIC and WIC tends to zero. In contaminated series the size of the test and of the information criteria increases with the outlier magnitude and the degree of persistence. SIC and WIC can clearly outperform the test in larger samples and larger outlier magnitudes. The power of the test and of the information criteria depends on the sample size and on the difference between the regimes. The more distinct the regimes and the larger the sample, the higher is the power. Additive outliers decrease the power in distinct regimes in small samples and in intermediate regimes in large samples, but increase the power in similar regimes. Due to their higher robustness in terms of size, information criteria are a valuable alternative to linearity tests in outlier contaminated time series.

JEL-Numbers: C15, C22

Keywords: Additive Outliers · Nonlinear Time Series · Information Criteria · Linearity Test · Monte Carlo

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

1 Introduction

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. (1988) becomes oversized in the presence of additive outliers (AOs), i.e. 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 (2015). 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 for 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 2 the definition of AR and SETAR models as well as the model framework of AOs are presented. In Section 3 the linearity test of Hansen (1999) and the information criteria are introduced. In Section

4 the simulation set-up and the simulation results are presented. Finally, Section 5 concludes.

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 autoregressions of order 1 (AR(1)) and two-regime self-exciting threshold autoregressive 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 $\varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2)$. For $t = 1, \ldots, n$ 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,

$$y_t = x_t(1 - \delta_t) + \zeta_t \delta_t. \tag{2.1}$$

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

$$\zeta_t = x_t + \zeta, \tag{2.2}$$

where ζ is the constant outlier magnitude. In order to model symmetric contaminations, the random variable δ_t takes the values -1 and 1 with a probability of $\pi/2$ respectively, and 0 with the probability $1-\pi$. The probability π is referred to as the outlier probability. Combining the definitions of Eq. (2.1) and (2.2), 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). Due to the fact that innovative outliers do not seriously deteriorate the performance of linearity tests, this paper only considers AOs (cf. van Dijk et al., 1999; Ahmad and Donayre, 2016).

The effect of an AO is illustrated in Figure 2.1. 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.

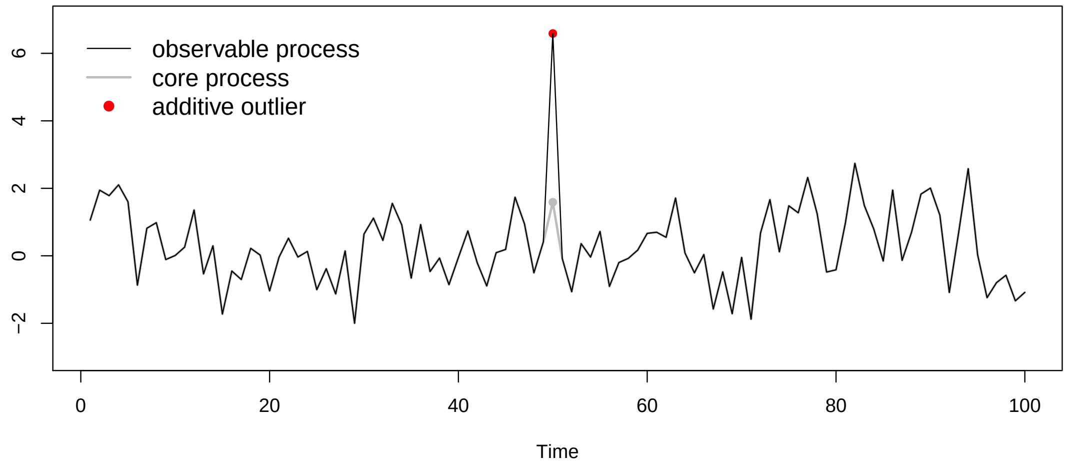


Figure 2.1: Effect of an Additive Outlier

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.

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 autoregressive 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 autoregressive 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

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

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 combination $(\hat{c}, \hat{d})$ that minimizes the residual sum of squares (RSS) is selected. The test statistic is then given by

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

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 should be simulated using bootstrap procedures. In this paper the bootstrap is designed as in Hansen (1999). The bootstrap is always conducted under the null hypothesis. So, bootstrap 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 bootstrap 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 bootstrap data the test is conducted and the test statistic is saved. The procedure is repeated 2000 times and the empirical distribution function of the bootstrap test statistics is calculated. If the test statistic of the original sample exceeds the $(1-\alpha)$-quantile of the bootstrap test statistics, the null hypothesis will be rejected.

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 (cf. Rinke and Sibbertsen, 2015),

$$SIC = n\log(\hat{\sigma}^2) + p\log(n),$$
$$WIC = n\log(\hat{\sigma}^2) + \frac{\left(2n(p+1)/(n-p-2)\right)^2 + \left(p\log(n)\right)^2}{2n(p+1)/(n-p-2) + p\log(n)}.$$

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 overall versions of the information criteria are calculated, i.e. the two regimes are not considered separately, but the goodness of fit and the number of parameters are assessed for the whole model (cf. Rinke and Sibbertsen, 2015). The threshold is included in the number of parameters of the SETAR model as an additional parameter (cf. Kapetanios, 2001; Rinke and Sibbertsen, 2015). The delay is not considered to be an additional parameter since it is fixed at 1 and does not have to be estimated. The estimator of the error term variance is given by

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

This estimator is unbiased and prevents overfitting (cf. McQuarrie et al., 1997; Rinke and Sibbertsen, 2015).

The information criteria are calculated for both models, AR(1) and SETAR(1,1), and the model which minimizes the respective information criterion is selected. So, again the computational effort is relatively high, because both models have to be fully specified. The estimation procedure is the same as for the linearity test.

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 $\pi = 0.05$. 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.

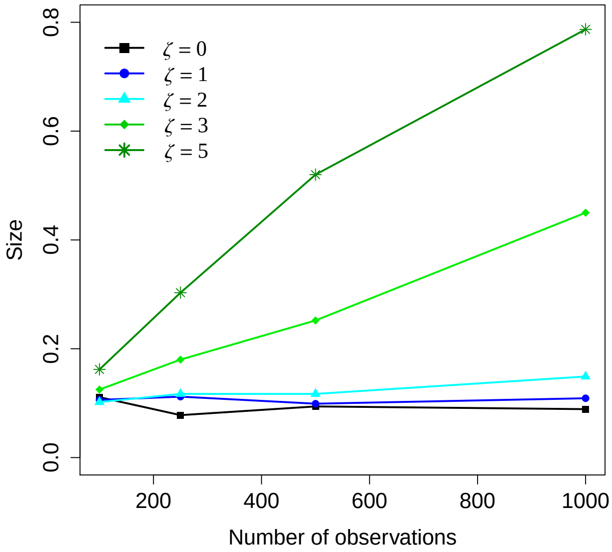
4.1 Size Properties

Table 4.1 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. The linearity test has good size properties independent of the sample size due to the application of the bootstrap. Only for a sample size of $n = 250$ and larger significance levels, the test appears to be slightly undersized. However, this is probably due to some data peculiarity. 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 (cf. also Fig. 4.1).

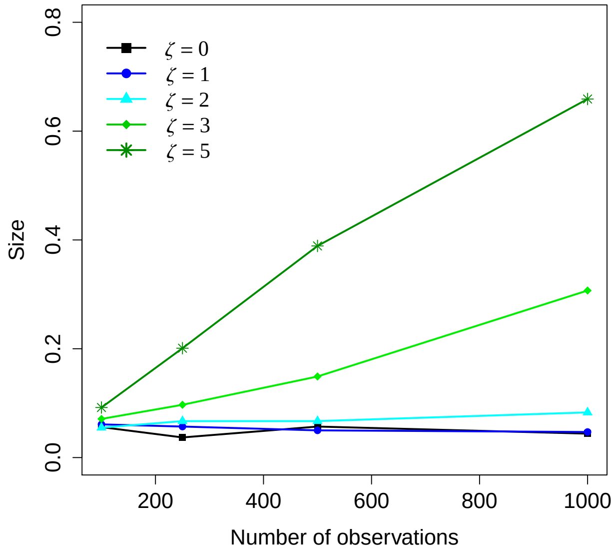
	Hansen (1999) Test			SIC	WIC
$n \backslash \alpha$	1%	5%	10%		
100	0.012	0.056	0.111	0.054	0.114
250	0.008	0.037	0.078	0.015	0.037
500	0.014	0.057	0.094	0.015	0.039
1000	0.010	0.044	0.089	0.004	0.014

Table 4.1: Size Properties of the Hansen (1999) Test and the Information Criteria. The DGP is an uncontaminated AR(1) process with $\phi = 0.25$.

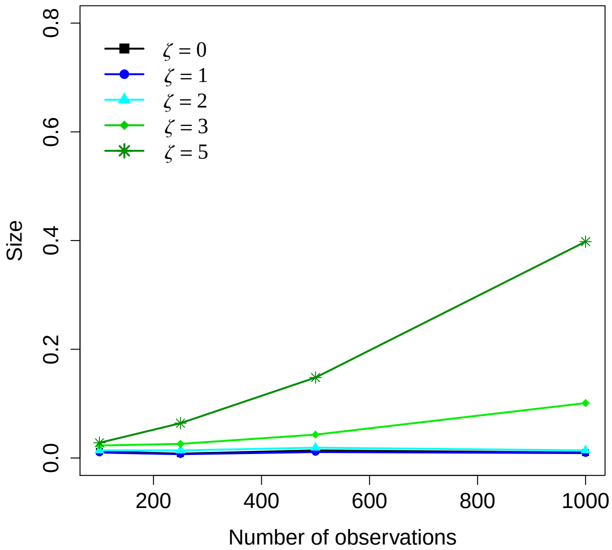
Figure 4.1 illustrates the effect of AOs with different outlier magnitudes $\zeta = \{0, 1, 2, 3, 5\}$ on the size of the linearity test and of the information criteria. In small samples $n = 100$ AOs only have a minor effect on the size of the linearity test and the information criteria. This is probably due to the fact that there are only a few outliers. The expected number of outliers equals $n \cdot \pi$. 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.



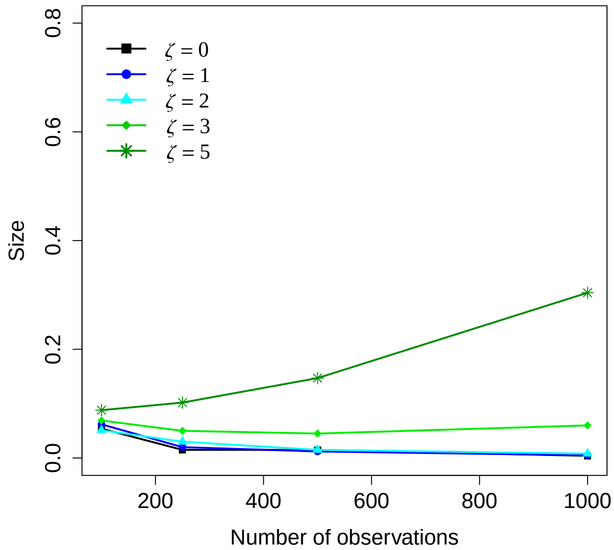
(a) $\alpha = 10\%$



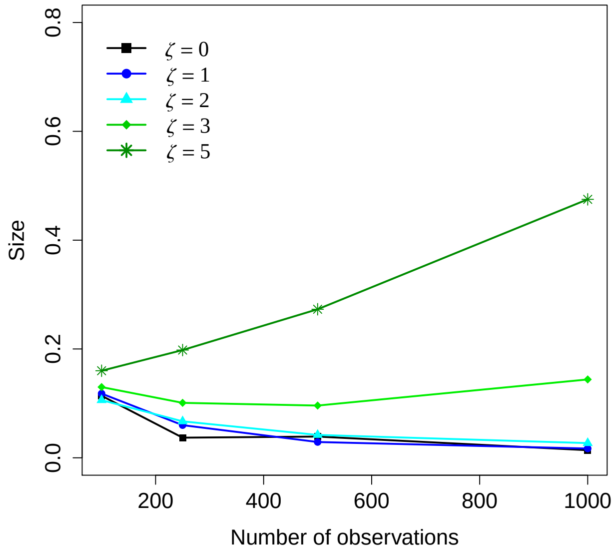
(b) $\alpha = 5\%$



(c) $\alpha = 1\%$



(d) SIC



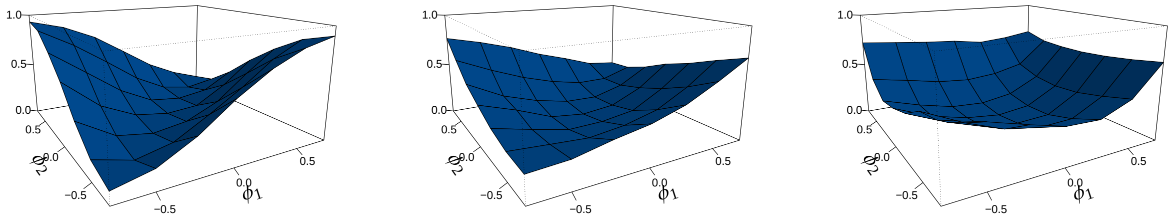
(e) WIC

Figure 4.1: 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 $\phi = 0.25$.

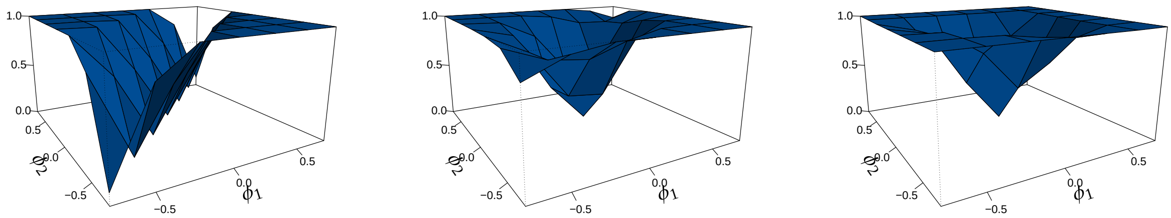
The effect of AOs on the size depends besides the outlier magnitude ζ and the sample size also on the degree of persistence. For $\phi = 0.75$ and $\zeta = 3$ the size of the test in the largest sample $n = 1000$ takes the values 0.554, 0.781, and 0.867 for the significance levels $\alpha = \{1\%, 5\%, 10\%\}$. SIC and WIC have a size of 0.454 and 0.621, respectively. Thus, the higher the degree of persistence, the higher are the size distortions. This coincides with the findings of Ahmad and Donayre (2016). The influence of the persistence on the size of contaminated processes can also be found in Figures 4.2 - 4.6.

4.2 Power Properties

The power properties of the linearity test and 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. This is illustrated in Figures 4.2 - 4.6. The more distinct the regimes, the higher is the power of the test and of the information criteria. For $\phi_1 = \phi_2$ the SETAR(1,1) model reduces to an AR(1) process. Thus, on the main diagonal the size is depicted and the influence of the degree of persistence in outlier contaminated series is pointed out again. For contaminated series the size increases, especially in highly persistent series. So, the plots of the contaminated series with $\zeta = 5$ turn upward at both ends of the diagonal. Therefore, the higher the absolute degree of persistence, the higher is the size distortion.

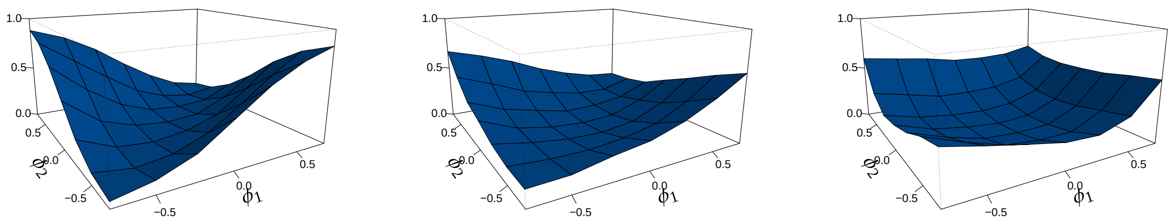


(a) $n = 100$ and $\zeta = 0$ **(b)** $n = 100$ and $\zeta = 3$ **(c)** $n = 100$ and $\zeta = 5$

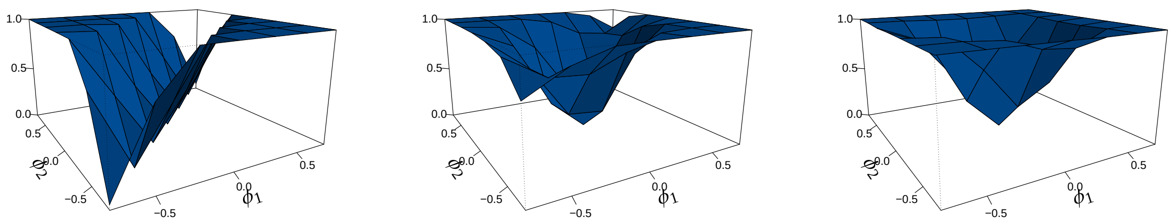


(d) $n = 1000$ and $\zeta = 0$ **(e)** $n = 1000$ and $\zeta = 3$ **(f)** $n = 1000$ and $\zeta = 5$

Figure 4.2: Power of the Hansen (1999) Test for $\alpha = 10\%$. The DGP is a SETAR(1,1) process.



(a) $n = 100$ and $\zeta = 0$ **(b)** $n = 100$ and $\zeta = 3$ **(c)** $n = 100$ and $\zeta = 5$

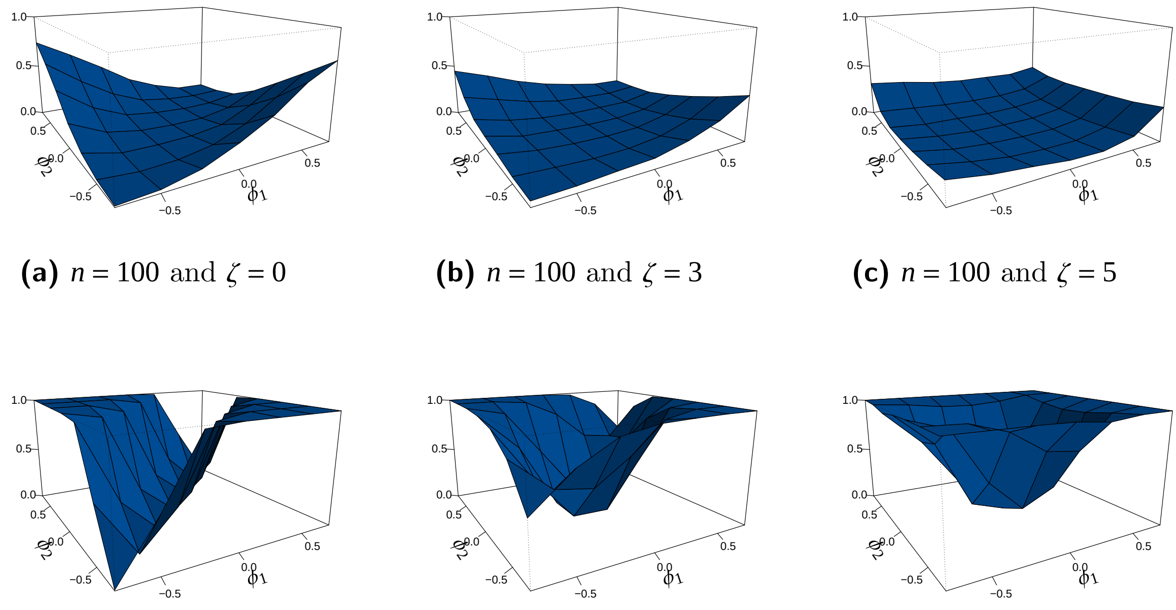


(d) $n = 1000$ and $\zeta = 0$ **(e)** $n = 1000$ and $\zeta = 3$ **(f)** $n = 1000$ and $\zeta = 5$

Figure 4.3: Power of the Hansen (1999) Test for $\alpha = 5\%$. The DGP is a SETAR(1,1) process.

Close to the diagonal the regimes are quite similar. In small samples the power of the test and of the information criteria in this scenario 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 in general, but also if the regimes are similar.

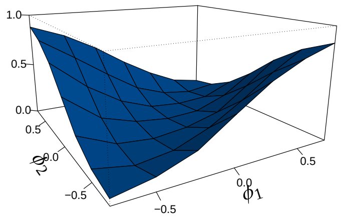
The introduction of AOs increases the size but slightly decreases the power in small samples for a nonlinear model with distinct regimes. For the larger number of observations the power converges to 1, but the size also increases enormously. The differentiation between AR(1) and SETAR(1,1), if the regimes are quite similar, is more reliable than for a small sample of a contaminated process. However, the increase in power is at least partly due to the overall increase of the selection frequency of the nonlinear model. So, there exists a trade-off between power gains and size distortions.



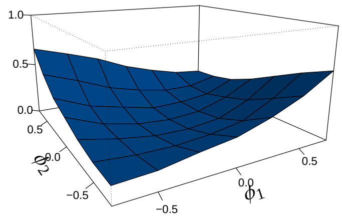
(a) $n = 100$ and $\zeta = 0$ **(b)** $n = 100$ and $\zeta = 3$ **(c)** $n = 100$ and $\zeta = 5$

(d) $n = 1000$ and $\zeta = 0$ **(e)** $n = 1000$ and $\zeta = 3$ **(f)** $n = 1000$ and $\zeta = 5$

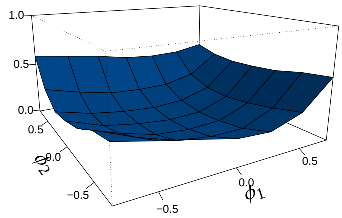
Figure 4.4: Power of the Hansen (1999) Test for $\alpha = 1\%$. The DGP is a SETAR(1,1) process.



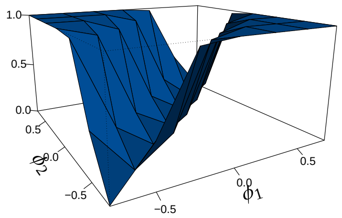
(a) $n = 100$ and $\zeta = 0$



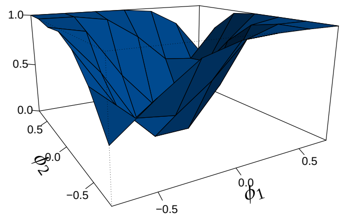
(b) $n = 100$ and $\zeta = 3$



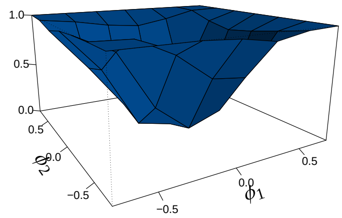
(c) $n = 100$ and $\zeta = 5$



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

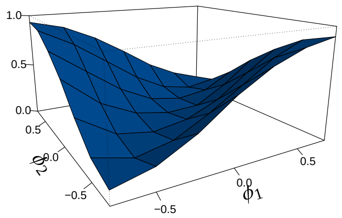


(e) $n = 1000$ and $\zeta = 3$

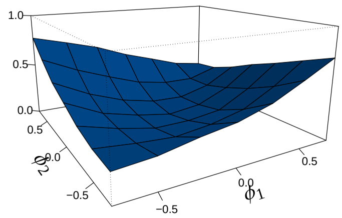


(f) $n = 1000$ and $\zeta = 5$

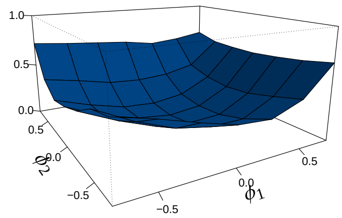
Figure 4.5: Power of the SIC. The DGP is a SETAR(1,1) process.



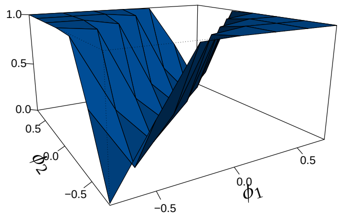
(a) $n = 100$ and $\zeta = 0$



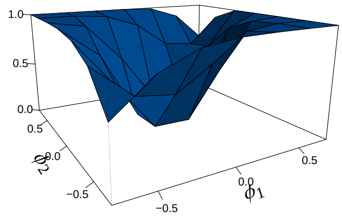
(b) $n = 100$ and $\zeta = 3$



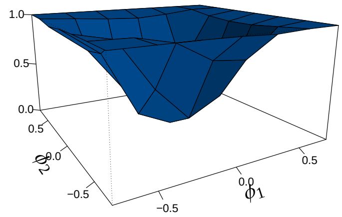
(c) $n = 100$ and $\zeta = 5$



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



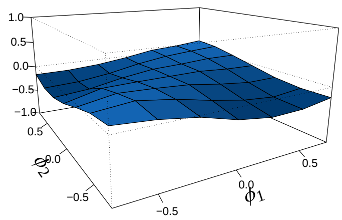
(e) $n = 1000$ and $\zeta = 3$



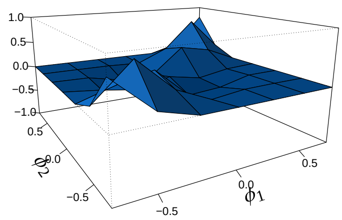
(f) $n = 1000$ and $\zeta = 5$

Figure 4.6: Power of the WIC. The DGP is a SETAR(1,1) process.

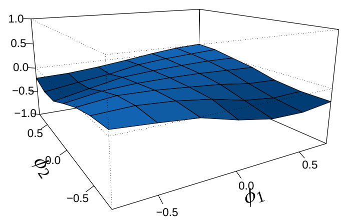
The change in power due to the introduction of AOs is visualized in Figure 4.7. The power for the outlier contaminated series with $\zeta = 3$ is compared to the uncontaminated series. Positive values indicate power gains, whereas negative values indicate power losses. The darker the plot, the smaller (more negative) is the change in power. So, for 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 the intermediate case if the outlier magnitude increases.



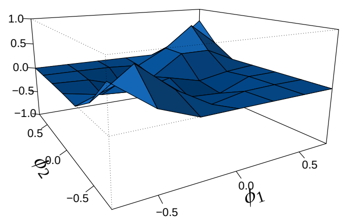
(a) $n = 100$ and $\alpha = 10\%$



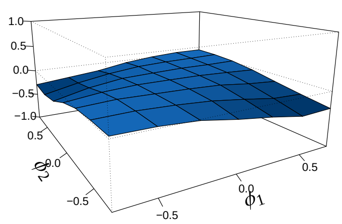
(b) $n = 1000$ and $\alpha = 10\%$



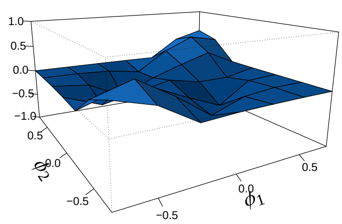
(c) $n = 100$ and $\alpha = 5\%$



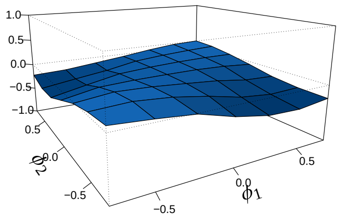
(d) $n = 1000$ and $\alpha = 5\%$



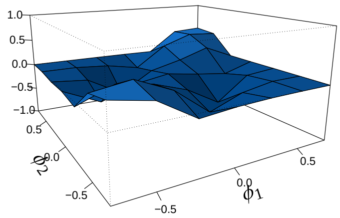
(e) $n = 100$ and $\alpha = 1\%$



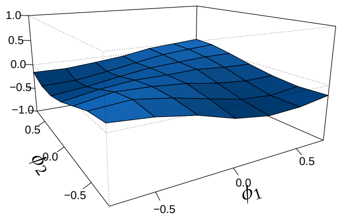
(f) $n = 1000$ and $\alpha = 1\%$



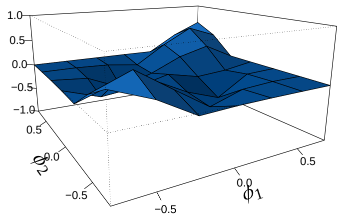
(g) $n = 100$ and SIC



(h) $n = 1000$ and SIC



(i) $n = 100$ and WIC



(j) $n = 1000$ and WIC

Figure 4.7: Change of Power for $\zeta = 3$ compared to $\zeta = 0$. The DGP is a SETAR(1,1) process.

In small samples the SIC and WIC perform like the linearity test at the 5% and 10% significance level. In larger samples their performance coincides with the 1% level. Thus, in terms of power the test cannot outperform the information criteria and vice versa.

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 $\zeta = \{3, 5\}$ and higher degrees of persistence ϕ the size increases. Especially in larger samples the results of the test 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 distortion in contaminated processes is 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. Therefore, the two information criteria are a valuable alternative to linearity tests due to their higher robustness against outlier contaminations in terms of size without considerable power losses.

References

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.

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

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

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.

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

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).

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

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

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

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

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

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

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. (2015). Information criteria for nonlinear time series models. *Studies in Nonlinear Dynamics & Econometrics*, forthcoming.

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

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.

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.