Information Criteria for Nonlinear Time Series Models*

Saskia Rinke and Philipp Sibbertsen[1]

Leibniz University Hannover

Abstract

In this paper the performance of different information criteria for simultaneous model class and lag order selection is evaluated using simulation studies. We focus on the ability of the criteria to distinguish linear and nonlinear models. In the simulation studies, we consider three different versions of the commonly known criteria AIC, SIC and AICc. In addition, we also assess the performance of WIC and evaluate the impact of the error term variance estimator. Our results confirm the findings of different authors that AIC and AICc favor nonlinear over linear models, whereas weighted versions of WIC and all versions of SIC are able to successfully distinguish linear and nonlinear models. However, the discrimination between different nonlinear model classes is more difficult. Nevertheless, the lag order selection is reliable. In general, information criteria involving the unbiased error term variance estimator overfit less and should be preferred to using the usual ML estimator of the error term variance.

JEL-Numbers: C15, C22

Keywords: Information Criteria · Nonlinear Time Series · Threshold Models · Monte Carlo

*The authors are grateful to the participants of the Statistische Woche in Hannover for helpful comments. Financial support by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.
[1]Corresponding Author. Leibniz University Hannover, School of Economics and Management, Institute of Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-Mail: sibbertsen@statistik.uni-hannover.de. Phone: +49-511-762-3783. Fax: +49-511-762-3923.

1 Introduction

In time series analysis the identification of a model that is able to appropriately describe special features of a given data set, like cyclical behavior or persistence of shocks in the series, is crucial. This is due to the fact that fitting a misspecified model to the data will lead to biased estimates and all further inference based on previous results, e.g. forecasting, will be misleading.

In order to identify the best fitting model there are two different strands of procedures in the literature, namely hypothesis testing and model selection using information criteria. In the context of linear time series models it is common practice to determine the lag order using information criteria. However, when nonlinear models are considered, the testing approach is preferred. So, instead of calculating information criteria for different models, linearity tests are applied (cf. Tong, 1990; Luukkonen et al., 1988a,b). In a first step a linear AR process is fitted to the data which lag length is determined using information criteria. Afterwards, this specification is tested against a nonlinear alternative (cf. Pitarakis, 2006; Luukkonen et al., 1988a). Pitarakis (2006) shows that the lag order selection can seriously influence the power properties of linearity tests. This is due to the fact, that the linear model will be misspecified if the true data generating process is actually nonlinear. Moreover, Luukkonen et al. (1988b) show that a linearity test designed to detect a certain kind of nonlinearity may also have power against other nonlinear models. Hence, the rejection of the null of a linear model may not tell which nonlinear model should be used to model the data.

Despite these drawbacks of testing, there are probably two reasons why the testing approach is preferred to model selection for nonlinear models. Firstly, it may not be clear how to calculate the value of an information criterion if a multiple regime model is fitted to the data. Information criteria can be easily calculated for single regimes, but then, these values have to be combined into one index in order to obtain one value for the whole model. Secondly, there exists no general rule if and how additional parameters of the nonlinear model have to be incorporated into the penalty terms of information criteria. Hence, the application of information criteria to nonlinear models may not result in the selection of optimal models (cf. Clements and Krolzig, 1998).

Nonetheless, in the literature there exist some examples of the application of information criteria to nonlinear time series models. Lag order selection is treated in Kapetanios (2001) for SETAR and MSAR models, in Smith et al. (2006) for MSAR models and in Tong (1983), Wong and Li (1998) and Li (1988) for SETAR models. Gonzalo and Pitarakis (2002) use information criteria to distinguish between linear AR models and single- and multiple-regime SETAR models. Further selection of the model class is considered among others in Psaradakis et al. (2009) and in Kapetanios (2001). Though,

in contrast to Psaradakis et al. (2009) and Gonzalo and Pitarakis (2002), Kapetanios (2001) incorporates the threshold parameters of the SETAR models into the penalty terms of information criteria. However, in most works the lag orders and other relevant parameters, like the delay parameter of SETAR and STAR models, are treated as given. Thus, the results are obtained under ideal conditions. In fact, if information criteria are applied to empirical data, the parameter values are unknown in advance and have to be estimated first. The resulting estimation errors can influence further calculations and hence, deteriorate the performance of the information criteria so that former results are not valid anymore. Therefore, in this work we apply different information criteria to select the optimal model class and the corresponding lag order and additional parameters simultaneously. The performance of the information criteria in different scenarios is assessed in several simulation studies. There, we will take three different versions of the respective criteria into account. Special focus will be on the fact, whether the criteria are able to successfully distinguish between linear and nonlinear time series models.
The rest of the paper is organized as follows. In Section 2 the nonlinear models which are considered in the simulation studies are explained. In Section 3 we shortly repeat the intuition of information criteria and introduce the four criteria we use. In Section 4 the simulation set-up and the simulation results are presented. Finally, Section 5 concludes.

2 Nonlinear Time Series Models

There exists a variety of different nonlinear time series models in the literature. In our simulation studies we focus on regime-switching models with switches in the mean equation. Hence, the class of ARCH (cf. Engle, 1982) and GARCH (cf. Bollerslev, 1986) models is not considered. The selection of ARCH/GARCH orders is treated e.g. in Hughes and King (2003) and Hughes et al. (2004). In the following the models used in the simulation studies are presented.

2.1 The SETAR Model

Self-exciting threshold autoregressive models were introduced in Tong and Lim (1980) and Tong (1983) (cf. also Tong, 1990). Since linear AR models are not able to capture certain nonlinear features of the data, but are easy to specify, the SETAR model is a natural extension of the linear model to the nonlinear case. SETAR models combine multiple piecewise linear regimes, which are separated by threshold parameters, into one model. A two regime SETAR model with p_1 and p_2 lags respectively can be written as

$$y_t = \begin{cases} \phi_{0_1} + \sum_{i=1}^{p_1} \phi_{i_1} y_{t-i} + \varepsilon_t, & \text{if } y_{t-d} > c\,; \\ \phi_{0_2} + \sum_{i=1}^{p_2} \phi_{i_1} y_{t-i} + \varepsilon_t, & \text{if } y_{t-d} \le c\,, \end{cases} \tag{2.1}$$

where $\varepsilon_t \sim iid(0, \sigma^2)$. The dependent variable y_t falls into the first regime which consists of an AR process with p_1 lags, if the threshold variable y_{t-d} exceeds the threshold c. Otherwise y_t falls into the second regime and follows an AR process with p_2 lags. In SETAR models th threshold variable is a lagged value of the dependent variable. The lag d is called the delay parameter and does not exceed the largest lag length (cf. Pitarakis, 2006). If in contrast the threshold variable is exogenous the SETAR model becomes a TAR model. Both models can be generalized to consist of m regimes. Then, there are m AR equations separated by $m-1$ threshold parameters.

2.2 The STAR Model

In contrast to the SETAR models, in STAR models the regime switches are not discrete jumps but smooth transitions. This implies that each observation does not lie in one single regime but is a weighted mixture of both regimes, where the transition function F_t attaches the weight to the respective regimes. So, the regimes cannot be clearly separated. In fact, there exists a continuum of regimes. The value of the transition function always lies in the unit interval. The following equation describes a STAR model with two regimes

$$y_t = \left(\phi_{0_1} + \sum_{i=1}^{p} \phi_{i_1} y_{t-i}\right)(1 - F_t(y_{t-d}, \gamma, c)) + \left(\phi_{0_2} + \sum_{i=1}^{p} \phi_{i_2} y_{t-i}\right) F_t(y_{t-d}, \gamma, c) + \varepsilon_t\,, \tag{2.2}$$

with $\varepsilon_t \sim iid(0, \sigma^2)$. Depending on the transition function F_t, the STAR model is a logistic STAR (LSTAR) model if

$$F_t(y_{t-d}, \gamma, c) = \frac{1}{1 + \exp(-\gamma(y_{t-d} - c))} \tag{2.3}$$

or an exponential STAR (ESTAR) model if

$$F_t(y_{t-d}, \gamma, c) = 1 - \exp\left(-\gamma(y_{t-d} - c)^2\right). \tag{2.4}$$

Like in SETAR models c denotes the threshold and d the delay parameter. The parameter γ regulates the speed of the transition between the regimes. For a value of $\gamma = 0$

there exists no regime shift, instead the model is linear, for $\gamma \to \infty$ the LSTAR model becomes a two-regime SETAR model, whereas the ESTAR model reduces to a linear model (cf. van Dijk et al., 2002; Teräsvirta, 1994; Luukkonen et al., 1988a). If the transition function F_t is the indicator function, Equation 2.2 describes a SETAR model. In Figure 2.1 the transition functions of the three regime-switching models are represented.

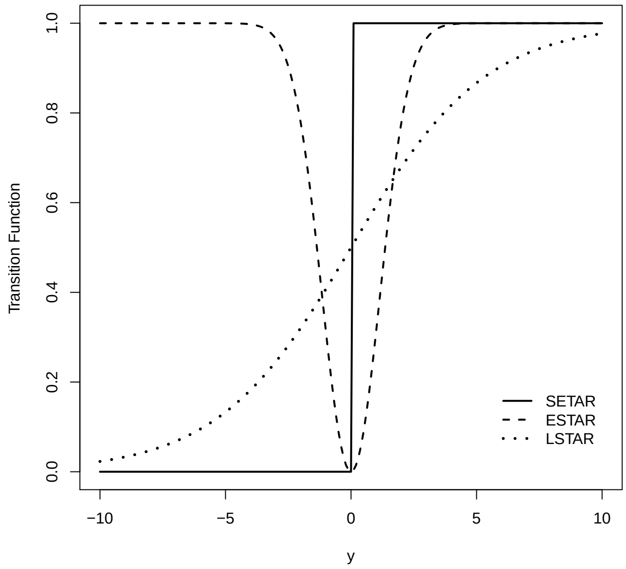


Figure 2.1: Transition Functions of SETAR and STAR models with $c = 0$ and $\gamma = 0.375$

In addition to SETAR and STAR models there also exists the class of Markov-switching autoregressive (MSAR) models (cf. Hamilton, 1989, 1994). However, in contrast to SETAR and STAR models, where the change of regime is governed by an endogenous variable, the regime shift in MSAR models is controlled by an exogenous, unobservable state variable. Due to this difference we do not consider MSAR models in our simulation studies but focus on the the ability of the information criteria to discriminate between linear and nonlinear models and to detect the correct form of the transition function.

3 Information Criteria

The idea of information criteria is to balance the goodness of fit and the complexity of a model using a loss function (cf. Wu and Sepulveda, 1998)

$$L = G(\hat{\sigma}^2) + P(n, p). \tag{3.1}$$

The first term of the loss function accounts for the goodness of fit and depends on an estimate of the unknown error term variance. The smaller the estimated variance of the error terms, the better is the model fit. The second term is the penalty term which depends on the sample size n and on the number of parameters p. Minimizing the loss function guarantees that if two models yield the same model fit, the model which contains fewer parameters is preferred. This is also known as the principle of parsimony (cf. Akaike, 1974; Schwarz, 1978).

3.1 The Traditional Information Criteria

Different choices of the penalty term yield different information criteria. We will shortly introduce the four information criteria we are using in our simulation studies.

The Akaike Information Criterion (AIC). The AIC introduced by Akaike (1974) as an estimate of the Kullback-Leibler information is probably the most commonly used information criterion

$$AIC = n\left(\log(\hat{\sigma}^2) + 1\right) + 2(p+1). \tag{3.2}$$

It has a rather weak penalty term which can result in overfitting in finite samples (cf. Hurvich and Tsai, 1989). This means that the selected model contains too many parameters.

The Schwarz Information Criterion (SIC). Using Bayes estimators, Schwarz (1978) derived another information criterion

$$SIC = n\log(\hat{\sigma}^2) + p\log(n). \tag{3.3}$$

The SIC has a stronger penalty term than the AIC in order to prevent overfitting. However, the SIC sometimes underfits. So, the selected model contains too few parameters. This is especially a problem in small samples.

The Corrected Akaike Information Criterion (AICc). Since the AIC is biased and therefore tends to overfit in finite samples, Hurvich and Tsai (1989) introduced a bias-corrected version of the AIC

$$AICc = n\log(\hat{\sigma}^2) + \frac{n(n+p)}{n-p-2}. \tag{3.4}$$

For small samples the AICc has a stronger penalty term than the AIC to solve the problem of overfitting. Asymptotically both versions are equivalent.

The Weighted Average Information Criterion (WIC). Wanting to combine the strengths of different criteria to obtain a criterion which performs well not depending on the sample size, Wu and Sepulveda (1998) introduce the weighted average information criterion

$$WIC = n\log(\hat{\sigma}^2) + \frac{(2n(p+1)/(n-p-2))^2 + (p\log(n))^2}{2n(p+1)/(n-p-2) + p\log(n)}. \tag{3.5}$$

The WIC is a weighted version of the AICc and the SIC. Setting the weights equal to the penalty terms of the respective criteria guarantees that for small samples the WIC behaves like AICc, which performs well in small samples. Besides, in large samples the WIC behaves like SIC, which performs well in large samples. These properties of WIC may be very valuable if we apply WIC seperately to the regimes e.g. of a SETAR model with a dominant regime. In this case, one regime contains significantly more observations than others. But due to the independence of the sample size, WIC should perform well in all regimes. We will further discuss and evaluate this point in Section 4.

3.2 The Versions of Information Criteria

As already mentioned in Section 1, it is not clear how to calculate the value of an information criterion for multiple-regime models. Since each information criterion depends on an estimated error term variance and the number of parameters of the model under consideration, it would be straightforward to estimate the error term variance using the residual sum of squares of the whole model and add the number of parameters of all regimes in order to obtain the number of parameters of the whole model. However, it is not clear whether information criteria maintain their optimality properties of linear specification, when they are applied to nonlinear models (cf. Clements and Krolzig, 1998). Thus, we also follow another approach. We consider two additional versions of information criteria, where we separate the models into their regimes. Due to the fact that the single regimes are linear, information criteria are supposed to select optimal lag orders of the regimes. Finally, the values of the information criteria of the single regimes have to be combined again into one global model value. A formal description is given below.

Equally Weighted Criteria. Following the approach of Tong (1983), for the first version we calculate the information criteria separately for each regime and then combine these values into one model information criterion, where each regime gets the equal weight

$$IC_{model} = \frac{1}{m}\sum_{i=1}^{m} IC_{reg.i}. \tag{3.6}$$

Regime Weighted Criteria. In the second version the weighting of the regimes is proportional to the dominance of the regime. The more dominant the respective regime is, the higher is the attached weight w_i

$$IC_{model} = \sum_{i=1}^{m} w_i\, IC_{reg.i} \qquad \text{with } w_i \in [0,1]. \tag{3.7}$$

For SETAR models the weights can simply be determined by the the number of observations that fall into the respective regime divided by the total number of observations. In STAR models, the value of the transition function can be used as a weight, since in STAR models, the observations do not fall in one regime only but are a weighted sum of both regimes.

Overall Model Criteria. In the third version the regimes are not considered separately. Instead, an overall criterion is calculated. Hence, the number of parameters in the penalty term equals the sum of lags of all regimes and the variance estimate is not calculated for each regime separately but for the whole model, i.e. all data points are considered in the computation (cf. Pitarakis, 2006).

The differentiation between these three versions is only meaningful for regime-switching models. Hence, for the linear AR model, which only consists of one single regime, all three versions are equivalent.

3.3 The Role of the Error Term Variance Estimator

All previously introduced information criteria depend on an estimate of the error term variance. Generally, the Maximum Likelihood estimator

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}, \tag{3.8}$$

where $\hat{\varepsilon}$ is the vector of residuals and n denotes the sample size, is used to calculate the values of the information criteria. However, $\hat{\sigma}^2$ is a biased estimator of the true error term variance. Therefore, McQuarrie et al. (1997) suggest to use

$$\tilde{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-p-1} \tag{3.9}$$

or the sample variance s^2 to evaluate different models using information criteria (cf. also Wen and Tu, 2001). According to McQuarrie et al. (1997) the information criteria involving $\tilde{\sigma}^2$ have a stronger penalty term and avoid overfitting.

4 Simulation Study

In the following simulation studies we evaluate the information criteria and their different versions presented in Section 3.

4.1 Data Generation

In order to identify factors that influence the performance of the information criteria we generate data from different linear and nonlinear models and vary the sample size, the persistence parameters, the degree of dominance and the number of regimes. The models of the basic set-up are tabulated in Tables A.1 - A.4 in the Appendix, where the ε_t form a Gaussian white noise process. Following Kapetanios (2001) for every sample size, we simulate 200 additional observations for each data generating process, which are discarded afterwards to avoid a starting value bias. All initial values are set to zero. The simulation results are based on 1000 replications. In our analysis we will primarily focus on single- and two-regime models, since the computational effort for multiple-regime models is rather high (cf. also Gonzalo and Pitarakis, 2002).

4.2 Model Estimation

After generating the data, we fit different linear and nonlinear models to the data, calculate the different versions of the information criteria and choose the model, which minimizes the respective information criterion. Throughout, we assume that the error terms ε_t are Gaussian. In order to minimize the computational effort, we set a maximum lag length of $p_{\max} = 4$ for the models fitted to the data. Thus, the largest AR model is an AR(4) and the largest regime switching models consist of an AR(4) specification in each regime. According to Luukkonen et al. (1988a) a lag order exceeding $p = 3$ is rather unlikely for a small sample size of $n = 50$, but probable for larger samples. In order to make the results among different sample sizes comparable, we choose $p_{\max} = 4$ for all sample sizes (cf. also Pitarakis, 2006; Tong and Lim, 1980). The effective sample size used to fit the models to the data and in all further calculations is thus $n - p_{\max}$ (cf. Wong and Li, 1998; Tong and Lim, 1980).

Parameter estimation is done by (conditional) least squares. For the threshold and delay parameter grids are constructed and the remaining parameters are estimated for each grid point. Following Hansen (1997) the grid of the threshold consists of the interval from the 15% to the 85% quantile of y_t. Disregarding the lower and upper 15% quantiles should guarantee that at least 15% of the data lie in each regime and hence, the number of observations in each regime is sufficient for persistence parameter estimation. The grid of the delay simply consists of integer values from 1 to $p_{\max}$. For each model the parameter combination which minimizes the residual sum of squares is selected and the corresponding values of the information criteria are calculated. In STAR models the grid search is done conditional on γ. According to Teräsvirta (1994) it is possible to standardize the exponent of the transition function and choose $\gamma = 1$ as a starting value. After determining all parameters the value of γ is adjusted by minimizing the residual sum of squares with respect to γ. For ESTAR models the exponent of the transition function is divided by the variance of y_t, whereas for LSTAR models the standard deviation of y_t is appropriate.

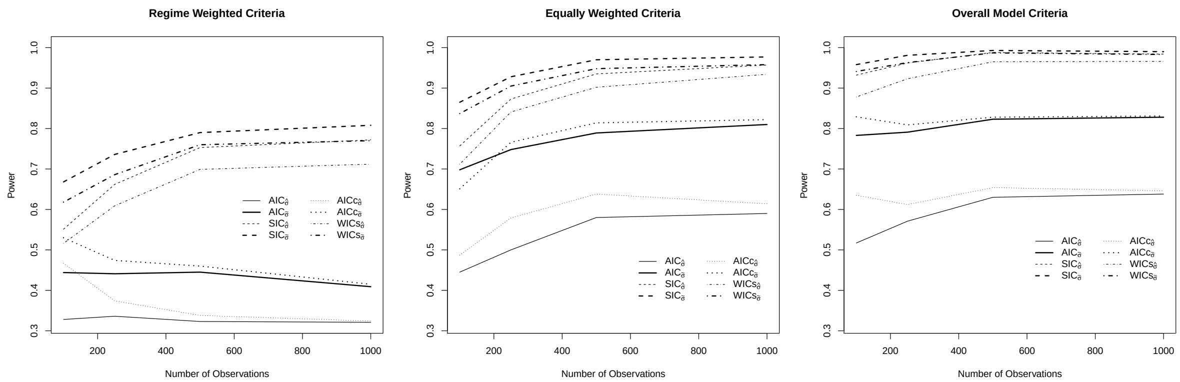
As already mentioned there exists no general rule if and how additional parameters of nonlinear models like the threshold and the delay should be incorporated into the penalty terms of information criteria. We decide to follow the approach of Kapetanios (2001) and add all additional parameters of the nonlinear models to the number of parameters. The intuition is that if the true DGP is nonlinear, then a nonlinear model will provide a better fit to the data. Hence, the value of the information criteria will decrease. However, the computational effort will increase. Since the information criteria are supposed to balance model fit and complexity, additional parameters are incorporated. As a result, the nonlinear model will only be selected if it provides a substantially better fit than a simpler model. For all three nonlinear models the additional parameters are the threshold parameter c as well as the delay parameter d. In STAR models we also consider the transition parameter γ as an additional parameter in the penalty term.

4.3 Simulation Results

In the following scenarios we assess the ability of the different information criteria to select the correct model. The Figures and Tables display the respective selection frequencies. In order to evaluate the performance let the power of an information criterion be defined as the relative frequency of selecting the correct model.

Lag Order Selection. This paragraph focuses on lag order selection within a certain model class. Hence, the following figures display the power of the information criteria when only the lag order (combination) within the true model class has to be determined. Although the assumption of knowing the correct model class will not be met if the information criteria are applied to real data, the power results will be helpful to correctly interpret the performance of the information criteria in further simulation studies. This is due to the fact, that if the criteria are not able to determine the correct lag order (combination) within the true model class, we cannot expect them to point to the correct model when also the model class has to be selected.

In Figure 4.1 the power of the three variants of information criteria are depicted when the true DGP is the LSTAR(1,1) model. It shows first characteristics of the information criteria. The regime weighted criteria perform worse than their equally weighted and

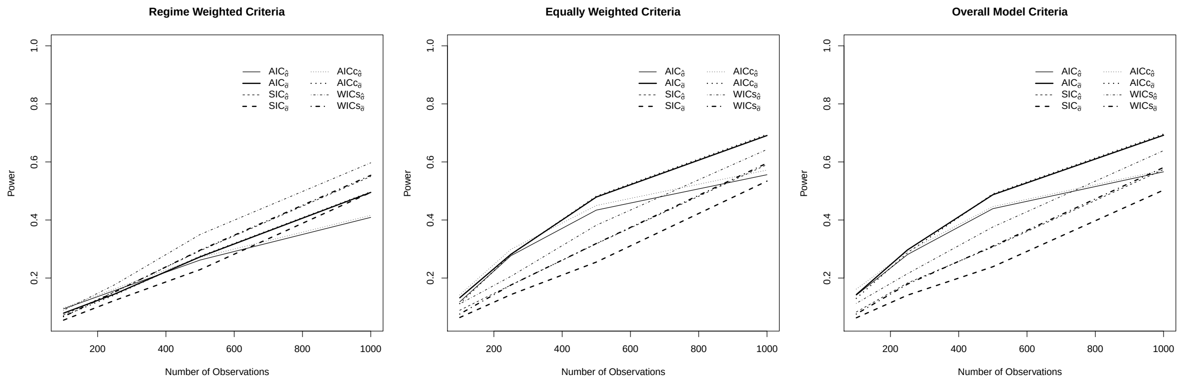


(a) Regime Weighted IC **(b)** Equally Weighted IC **(c)** Overall IC

Figure 4.1: Power of the Information Criteria (IC) for LSTAR Models (LSTAR(1,1))

overall counterparts. All versions of AIC and AICc cannot select the correct model with a probability approaching 1 when the sample size increases. This is due to the fact, that AIC and AICc are not consistent information criteria (cf. Shibata, 1986). Instead of the correct lag order combination AIC and AICc tend to overfit and choose larger combinations. However, AICc performs better than AIC in small and moderate samples (cf. also Wong and Li, 1998). Using the unbiased error term variance estimator $\tilde{\sigma}^2$ reduces the probability of overfitting. So, the versions of AIC and AICc involving $\tilde{\sigma}^2$ improve up to 20 percentage points (cf. Figures 4.1b and 4.1c).

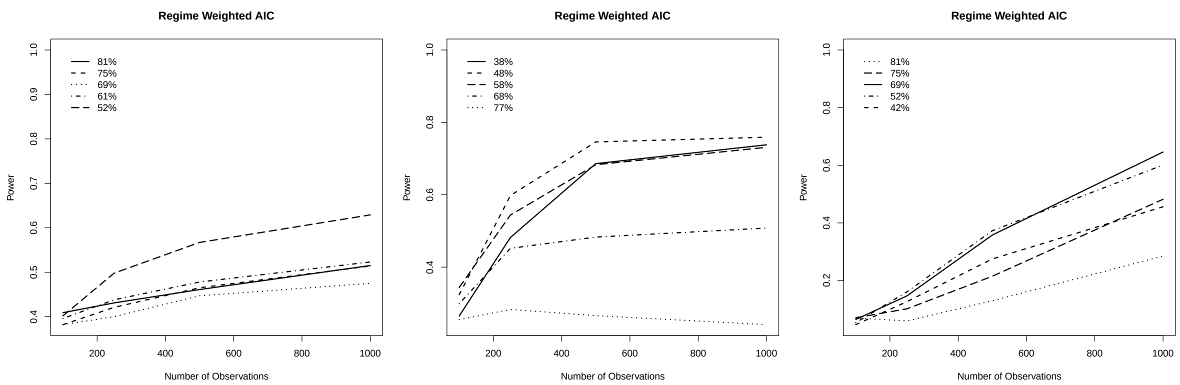
In general, all information criteria perform better in small models. With increasing lag order combinations, there is a tendency to underestimate the lag order of one regime independent of the error term variance estimator. Naturally, AIC and AICc outperform SIC and WIC in these cases due to their tendency to overfit. However, underfitting can occur due to weakly identifiable models (cf. McQuarrie et al., 1997). If the largest lag order only has a minor influence, it may be neglected and a smaller model is preferred like for the SETAR(1,3) model (cf. Figure 4.2). Instead of a SETAR(1,3) model a



(a) Regime Weighted IC **(b)** Equally Weighted IC **(c)** Overall IC

Figure 4.2: Power of the Information Criteria (IC) for SETAR Models (SETAR(1,3))

SETAR(1,2) is preferred in small and moderate samples. Equally weighted and overall AIC and AICc (especially with the unbiased error term variance) outperform SIC and WIC. Regime weighted information criteria perform poorly. The idea to include the regime weighted information criteria and also the WIC into the simulation study is that they may lead to better results if one regime is dominant in the model. In general, the observations do not fall equally in the regimes, i.e. one regime is always more dominant than another. Therefore, we consider three modified DGPs. From the SETAR(1,1), the SETAR(2,1) and the SETAR(3,2) model we generate data with five different thresholds, resulting in a first regime with a share of observations varying from 38% to 81%. Afterwards, we fit SETAR models to the data and assess the effect of dominant regimes. In the first two panels of Figure 4.3 we see that the power of the regime weighted AIC with the unbiased error term variance estimator $\tilde{\sigma}^2$ is highest if the first regime is not more dominant than the second. So, the best performance is achieved if 52% of the observations fall into the first regime for the SETAR(1,1)and 48% for the SETAR(2,1), respectively. But in fact, the size of the effect is quite different. Although for the SETAR(1,1) the fair separation of regimes is definitely the best, the difference between the best and the worst separation (with 69% of the observations in the first regime) only amounts to approximately 15 percentage points in large samples. In contrast, for the SETAR(2,1) model the fair separation yields similar results as if 58% and asymptotically similar results as if 38% of the observations lie in the first regime. Nevertheless, the difference between the best and the worst separation (with 77% of the observations in the first regime) amounts to 50 percentage points in large samples. In small samples, the difference between dominant and fair regimes is rather small. The comparison points out another fact: The worst separation is not always the one with one very dominant regime (cf. Fig. 4.3a). These considerations are valid for all regime weighted information criteria. Though, the effects are more pronounced for AIC and AICc. The equally weighted and overall information criteria are more independent from the dominance of the regimes. So asymptotically all separations yield the same power results. For moderate samples the effect is mostly pronounced.



(a) SETAR(1,1) **(b)** SETAR(2,1) **(c)** SETAR(3,2)

Figure 4.3: Power of the Information Criteria in Dominant Regimes: Regime Weighted AIC with Unbiased Error Term Variance Estimator

Considering Figure 4.3c, we see that the separations with 52% and 69% of the observations in the first regime offer the best power results if the regime weighted AIC is applied. The AIC performs worst if the first regime is most dominant (81% of the observations). However, for this model, other regime weighted criteria behave differently. So, SIC and WIC perform best if 69% or 75% of the observations fall into the first regime. This is similar to the equally weighted and overall information criteria. A dominant first regime leads to better power results than a fair separation (52%) or a dominant second regime (42%). Hence, there is no general proposition that the performance of the regime weighted information criteria is proportional to the degree of dominance of one regime, e.g. the one with a larger lag order. Instead, the behavior appears to be DGP specific.

Discriminating Linear and Nonlinear Models. In this paragraph we assess the performance of the information criteria for selecting between the linear and the nonlinear model. Again, we only consider the correct nonlinear model class. In the following Tables the blue row indicates the correct model, whereas the bold numbers mark the models with the highest selection frequency for the respective criterion. We only give an extract of the whole Tables for n=100 and n=1000. The results for n=250 and n=500 can be found in the appendix.

In Table 4.1 the selection frequencies are given when the data is generated by a SETAR(1,1) process with 100 observations. The first interesting point is that weighted versions of AIC and AICc never select a linear model. Although it is the correct decision in this setting, it is shown e.g. in the next paragraph that this behavior is spurious. This fact is also pointed out by Gonzalo and Pitarakis (2002) and Pitarakis (2006). In contrast, SIC and WIC have the tendency to select the linear model (cf. Psaradakis et al., 2009). However, taking into account the results for n=1000 in Table 4.2, it becomes obvious that this is a problem in small samples due to the fact that for the increased sample size all criteria prefer the correct model. Comparing these results with Tables 4.3 and 4.4, it becomes evident that the tendency to select linear models also depends on the true model structure. So, in symmetric models, where the lag order is equivalent in both regimes, SIC and WIC tend to prefer the linear over the nonlinear model. If the lag orders differ among regimes, it becomes easier to detect the nonlinearity. However, the overall SIC and WIC still tend to select the linear model (cf. Table 4.3).

Considering the results for the SETAR(2,3) model in Table 4.4, it is striking that the regime weighted information criteria vary a lot more than their equally weighted and overall counterparts, which focus on one model for a large sample size. This fact also partly carries over to AIC and AICc. The respective versions of SIC and WIC concentrate more on one model. So, for the equally weighted versions using the common ML error term variance estimator the selection frequency of the SETAR(2,2) model when n=1000 is about 40 percentage points higher for SIC and WIC than for AIC and still about 30 percentage points higher than for AICc (cf. Table 4.4).

The Tables presented in this paragraph pick up the problem of underfitting already mentioned in the previous paragraph. Especially larger lag order combinations are not estimated correctly. However, this might be due to identification problems.

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	**0.471**	**0.494**	0.000	0.000	**0.398**	**0.436**
SETAR(1,1)	**0.263**	**0.382**	0.263	0.301	**0.334**	**0.449**	0.273	0.330
SETAR(1,2)	0.118	0.140	0.068	0.071	0.126	0.134	0.080	0.080
SETAR(1,3)	0.077	0.073	0.043	0.029	0.080	0.064	0.052	0.030
SETAR(1,4)	0.090	0.067	0.029	0.014	0.071	0.053	0.026	0.016
SETAR(2,1)	0.085	0.091	0.048	0.046	0.095	0.093	0.061	0.054
SETAR(3,1)	0.088	0.074	0.024	0.018	0.085	0.068	0.032	0.021
SETAR(4,1)	0.072	0.057	0.023	0.011	0.057	0.048	0.019	0.011

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	**0.492**	**0.516**	0.000	0.000	**0.461**	**0.488**
SETAR(1,1)	**0.252**	**0.525**	0.299	0.380	**0.473**	**0.693**	0.319	0.405
SETAR(1,2)	0.115	0.114	0.049	0.031	0.124	0.107	0.052	0.039
SETAR(1,3)	0.093	0.054	0.026	0.013	0.059	0.020	0.022	0.007
SETAR(1,4)	0.105	0.045	0.021	0.007	0.045	0.018	0.018	0.005
SETAR(2,1)	0.111	0.105	0.043	0.027	0.124	0.078	0.040	0.029
SETAR(3,1)	0.092	0.066	0.022	0.011	0.076	0.040	0.023	0.009
SETAR(4,1)	0.066	0.035	0.017	0.003	0.035	0.017	0.009	0.001

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.050	0.178	**0.806**	**0.863**	0.100	0.238	**0.591**	**0.710**
SETAR(1,1)	**0.272**	**0.479**	0.143	0.108	**0.342**	**0.491**	0.272	0.232
SETAR(1,2)	0.105	0.086	0.012	0.008	0.103	0.079	0.030	0.014
SETAR(1,3)	0.088	0.037	0.003	0.001	0.081	0.028	0.012	0.004
SETAR(1,4)	0.071	0.028	0.000	0.000	0.058	0.016	0.006	0.000
SETAR(2,1)	0.106	0.072	0.008	0.006	0.099	0.063	0.032	0.014
SETAR(3,1)	0.083	0.041	0.000	0.000	0.068	0.028	0.011	0.001
SETAR(4,1)	0.052	0.028	0.001	0.000	0.043	0.020	0.007	0.001

(c) Overall Information Criteria

Table 4.1: Selection Frequencies of the Information Criteria: AR vs. SETAR models for SETAR(1,1) DGP with n=100

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.086	0.088	0.000	0.000	0.063	0.064
SETAR(1,1)	**0.360**	**0.475**	**0.637**	**0.668**	**0.369**	**0.477**	**0.622**	**0.653**
SETAR(1,2)	0.150	0.134	0.105	0.094	0.147	0.134	0.108	0.106
SETAR(1,3)	0.068	0.065	0.062	0.058	0.069	0.064	0.067	0.062
SETAR(1,4)	0.074	0.060	0.035	0.031	0.070	0.060	0.044	0.037
SETAR(2,1)	0.103	0.088	0.035	0.030	0.103	0.090	0.043	0.037
SETAR(3,1)	0.062	0.057	0.014	0.011	0.063	0.056	0.018	0.014
SETAR(4,1)	0.049	0.034	0.006	0.006	0.048	0.033	0.007	0.006

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.106	0.106	0.000	0.000	0.090	0.091
SETAR(1,1)	**0.465**	**0.703**	**0.838**	**0.868**	**0.483**	**0.710**	**0.813**	**0.849**
SETAR(1,2)	0.130	0.095	0.030	0.016	0.123	0.093	0.047	0.030
SETAR(1,3)	0.066	0.040	0.002	0.002	0.067	0.037	0.009	0.002
SETAR(1,4)	0.053	0.020	0.002	0.000	0.049	0.020	0.003	0.002
SETAR(2,1)	0.105	0.076	0.015	0.003	0.106	0.077	0.026	0.018
SETAR(3,1)	0.056	0.030	0.004	0.002	0.056	0.029	0.006	0.004
SETAR(4,1)	0.053	0.016	0.000	0.000	0.050	0.015	0.000	0.000

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.012	0.019	0.000	0.000	0.002	0.003
SETAR(1,1)	**0.501**	**0.728**	**0.952**	**0.970**	**0.506**	**0.734**	**0.935**	**0.958**
SETAR(1,2)	0.110	0.081	0.020	0.008	0.111	0.080	0.029	0.019
SETAR(1,3)	0.064	0.039	0.001	0.000	0.064	0.037	0.003	0.002
SETAR(1,4)	0.047	0.023	0.000	0.000	0.047	0.022	0.001	0.000
SETAR(2,1)	0.101	0.069	0.011	0.002	0.100	0.070	0.024	0.014
SETAR(3,1)	0.057	0.028	0.004	0.001	0.056	0.028	0.005	0.004
SETAR(4,1)	0.047	0.015	0.000	0.000	0.048	0.014	0.000	0.000

(c) Overall Information Criteria

Table 4.2: Selection Frequencies of the Information Criteria: AR vs. SETAR models for SETAR(1,1) DGP with n=1000

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.166	0.190	0.000	0.000	0.121	0.149
AR(2)	0.000	0.000	0.094	0.079	0.000	0.000	0.111	0.086
SETAR(1,1)	0.157	**0.242**	0.171	**0.230**	**0.211**	**0.290**	0.162	**0.223**
SETAR(1,2)	**0.166**	0.202	**0.184**	0.202	0.186	0.221	**0.188**	0.209
SETAR(1,3)	0.083	0.077	0.054	0.041	0.079	0.068	0.051	0.042
SETAR(2,1)	0.082	0.100	0.071	0.073	0.109	0.111	0.073	0.090
SETAR(2,2)	0.082	0.075	0.090	0.076	0.085	0.077	0.103	0.081
SETAR(2,3)	0.051	0.039	0.028	0.019	0.036	0.027	0.037	0.022

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.166	0.199	0.000	0.000	0.136	0.172
AR(2)	0.000	0.000	0.109	0.091	0.000	0.000	0.134	0.109
SETAR(1,1)	0.077	0.210	0.176	**0.258**	0.182	**0.322**	0.169	**0.264**
SETAR(1,2)	0.142	**0.238**	**0.182**	0.195	**0.214**	0.262	**0.195**	0.205
SETAR(1,3)	0.065	0.056	0.046	0.031	0.056	0.039	0.044	0.026
SETAR(2,1)	0.083	0.126	0.077	0.076	0.137	0.143	0.084	0.084
SETAR(2,2)	**0.148**	0.149	0.108	0.090	0.149	0.115	0.112	0.082
SETAR(2,3)	0.088	0.040	0.027	0.014	0.044	0.020	0.020	0.009

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.009	0.040	**0.351**	**0.463**	0.014	0.059	**0.199**	**0.293**
AR(2)	0.016	0.035	0.237	0.224	0.026	0.056	0.188	0.195
SETAR(1,1)	0.075	0.187	0.108	0.105	0.102	0.210	0.143	0.171
SETAR(1,2)	**0.163**	**0.239**	0.149	0.117	**0.184**	**0.244**	0.191	0.178
SETAR(1,3)	0.063	0.049	0.016	0.010	0.059	0.046	0.027	0.017
SETAR(2,1)	0.073	0.099	0.044	0.033	0.090	0.100	0.069	0.053
SETAR(2,2)	0.161	0.146	0.057	0.034	0.169	0.142	0.104	0.061
SETAR(2,3)	0.083	0.047	0.005	0.001	0.075	0.025	0.012	0.004

(c) Overall Information Criteria

Table 4.3: Selection Frequencies of the Information Criteria:
AR vs. SETAR models for SETAR(2,3) DGP with n=100

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SETAR(1,1)	**0.230**	0.237	0.005	0.008	**0.230**	0.238	0.004	0.005
SETAR(1,2)	0.225	**0.247**	0.012	0.015	0.228	**0.249**	0.010	0.012
SETAR(1,3)	0.050	0.042	0.000	0.000	0.049	0.040	0.000	0.000
SETAR(2,1)	0.031	0.034	0.130	0.141	0.031	0.034	0.118	0.130
SETAR(2,2)	0.170	0.257	**0.691**	**0.709**	0.179	0.260	**0.662**	**0.697**
SETAR(2,3)	0.083	0.067	0.118	0.094	0.081	0.065	0.137	0.112

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001
SETAR(1,1)	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001
SETAR(1,2)	0.000	0.000	0.006	0.008	0.000	0.000	0.003	0.006
SETAR(1,3)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SETAR(2,1)	0.001	0.003	0.022	0.037	0.001	0.005	0.013	0.023
SETAR(2,2)	**0.481**	**0.686**	**0.846**	**0.879**	**0.509**	**0.695**	**0.812**	**0.856**
SETAR(2,3)	0.215	0.173	0.103	0.067	0.205	0.169	0.133	0.094

(b) Equally Weighted Information Criteria

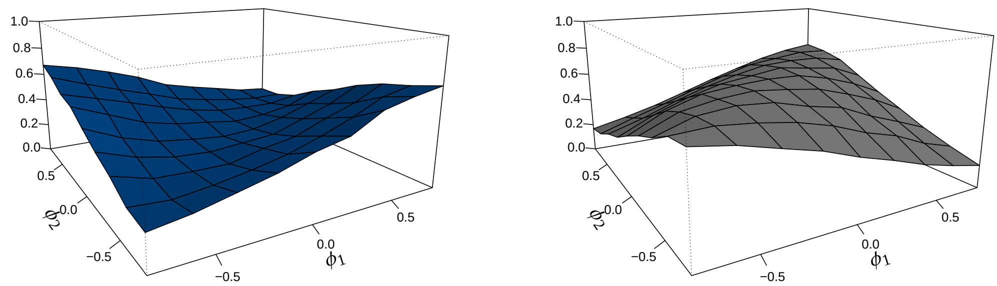
	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SETAR(1,1)	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001
SETAR(1,2)	0.000	0.000	0.011	0.013	0.000	0.000	0.006	0.011
SETAR(1,3)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SETAR(2,1)	0.001	0.002	0.026	0.041	0.001	0.002	0.015	0.024
SETAR(2,2)	**0.459**	**0.673**	**0.900**	**0.899**	**0.468**	**0.678**	**0.872**	**0.894**
SETAR(2,3)	0.233	0.185	0.053	0.040	0.235	0.182	0.083	0.058

(c) Overall Information Criteria

Table 4.4: Selection Frequencies of the Information Criteria:
AR vs. SETAR models for SETAR(2,3) DGP with n=1000

The previous remarks on SETAR models are also valid for the selection between AR and ESTAR and LSTAR models, respectively. However, the differentiation between linear and nonlinear models is slightly better for LSTAR than for ESTAR models. This is due to the asymptotic behavior concerning the transition parameter γ. ESTAR models converge to linear models for both extremes $\gamma \to 0$ and $\gamma \to \infty$, whereas LSTAR models only become linear for $\gamma \to 0$. Thus, the parameter estimate of the transition variable plays an important role. This will be further discussed in the context of discrimination between nonlinear models.

Effects of the Persistence Parameters. According to Psaradakis et al. (2009) the performance of the information criteria to select between the linear and the nonlinear model is better when the persistence parameters among the regimes differ. The following Figures display this fact and the dependence on the sample size. The true DGP is a SETAR(1,1) model with persistence parameters varying from -0.8 to 0.8 by 0.2. Since the error terms of both regimes are iid normally distributed, the SETAR(1,1) reduces to an AR(1) model if the persistence parameters are equal.

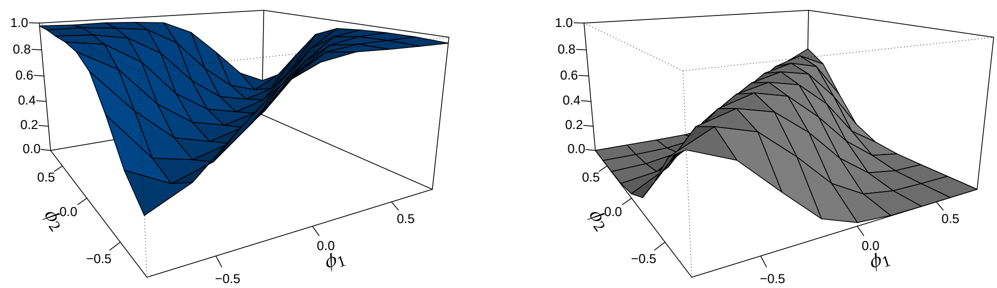


(a) SETAR(1,1) **(b)** AR(1)

Figure 4.4: Selection Frequencies for Equally Weighted SIC with $\tilde{\sigma}^2$ and n=100

For a small sample size of $n = 100$ the selection frequency of a SETAR(1,1) model increases with the distance between the persistence parameters (cf. Figure 4.4a). This confirms the findings of Psaradakis et al. (2009) that the differentiation between linear and nonlinear models becomes easier the more the regimes differ. On the diagonal the persistence parameters are equal and the model reduces to an AR(1) process. On this diagonal the selection frequency for the SETAR(1,1) is lowest. However, the selection frequency of the SETAR(1,1) model is already relatively low when the regimes are not equal but quite similar. This is confirmed by Figure 4.4b which depicts the respective selection frequency of an AR(1) model. On the diagonal the true model is actually the AR(1) model and there the selection frequency is the highest. Close to the diagonal, the selection frequency is still rather high. This implies that the linear model is preferred to the nonlinear model. For distinct regimes the AR(1) model is clearly inferior to the SETAR(1,1) model.

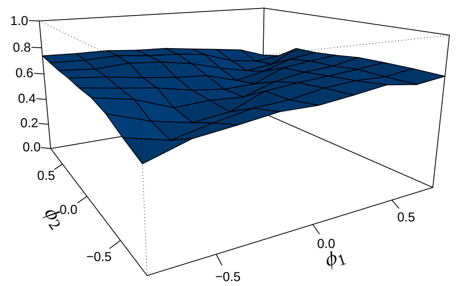
With an increasing sample size the differentiation between linear and nonlinear models is more reliable. In Figure 4.5a the selection frequency of the SETAR(1,1) model approaches 1 if the regimes are distinct. The more similar the regimes become, the lower is the selection frequency. Again, it is minimal on the diagonal where the model reduces to the linear case. Additionally, in Figure 4.5b the respective selection frequencies of the AR(1) model are presented. For distinct persistence parameters the linear model is never selected. The more similar the regimes become, the higher is the selection frequency. Nevertheless, the regimes have to be very similar, otherwise the nonlinear model is superior to the linear model. Hence, there are more correct selections if the number of observations increases. All versions of SIC and WIC have these properties.



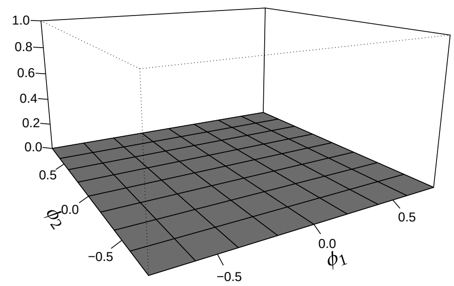
(a) SETAR(1,1)　　**(b)** AR(1)

Figure 4.5: Selection Frequency for Equally Weighted SIC with $\tilde{\sigma}^2$ and n=1000

The overall versions of AIC and AICc are able to detect linearity but especially in small samples have a tendency to prefer the nonlinear over the linear model. The weighted versions of AIC and AICc cannot detect linearity. They spuriously select the nonlinear model and never the linear model independent of the sample size and the distance between regimes. Figure 4.6 illustrates the spurious behavior of the equally weighted AIC already mentioned in the previous paragraph.



(a) SETAR(1,1)



(b) AR(1)

Figure 4.6: Selection Frequency for Equally Weighted AIC with $\tilde{\sigma}^2$ and n=1000

Discriminating Nonlinear Models. So far we have only considered the selection between linear and nonlinear models, focussing on one nonlinear model class. Now, the collection of models includes also other nonlinear model classes in order to assess the ability of the information criteria to determine the form of the transition function. In a first step we only consider the LSTAR and the ESTAR models (cf. Tab. 4.5 - 4.8). Our results point out that there are two important factors that influence the information criteria. The first one is the sample size: As already pointed out by Psaradakis et al. (2009), it is difficult to determine the switching mechanism when only a small number of observations is available. Moreover, the transition parameter γ plays an important role. The selection results are better for a small value of $\gamma = 1$ than for $\gamma = 20$. In the small sample the equally weighted and the overall criteria favor the ESTAR over the LSTAR model (cf. Tables 4.5 and 4.7). Though, the selected lag orders are appropriate. In fact, the equally weighted and the overall criteria tend to a spurious selection of the ESTAR model class for larger γ, whereas the regime weighted criteria favor LSTAR models. However, the lag order selection is always similar to the lag order combination preferred if only the correct model class is considered. With an increasing sample size both lag order and model class selection improve for all versions of information criteria (cf. Tables 4.6 and 4.8). Nevertheless, the selection frequencies of the correct model are higher if $\gamma = 1$.

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.010	0.011	0.066	0.089	0.008	0.011	0.103	0.123
ESTAR(2,1)	0.006	0.012	0.039	0.051	0.020	0.023	0.079	0.091
ESTAR(2,2)	0.026	0.032	0.183	0.156	0.023	0.020	0.160	0.147
ESTAR(2,3)	0.007	0.007	0.016	0.006	0.009	0.007	0.021	0.011
LSTAR(1,1)	0.131	0.193	0.092	0.134	**0.238**	**0.305**	0.063	0.101
LSTAR(2,1)	0.140	0.191	0.109	0.139	0.204	0.229	0.079	0.092
LSTAR(2,2)	**0.200**	**0.237**	**0.276**	**0.276**	0.195	0.201	**0.289**	**0.300**

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.064	0.108	0.101	0.146	0.094	0.125	0.124	0.158
ESTAR(2,1)	0.021	0.041	0.047	0.066	0.091	0.125	0.078	0.106
ESTAR(2,2)	0.225	0.274	0.295	0.277	0.172	0.171	0.232	0.206
ESTAR(2,3)	0.035	0.021	0.024	0.010	0.027	0.018	0.019	0.010
LSTAR(1,1)	0.004	0.022	0.031	0.048	0.008	0.019	0.025	0.042
LSTAR(2,1)	0.017	0.029	0.046	0.071	0.037	0.058	0.038	0.058
LSTAR(2,2)	**0.237**	**0.314**	**0.271**	**0.285**	**0.317**	**0.347**	**0.345**	**0.346**

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.059	0.114	0.176	0.212	0.082	0.126	0.143	0.187
ESTAR(2,1)	0.017	0.043	0.065	0.076	0.031	0.049	0.054	0.068
ESTAR(2,2)	0.247	0.285	0.266	0.230	0.279	0.292	0.284	0.261
ESTAR(2,3)	0.032	0.025	0.006	0.005	0.028	0.016	0.015	0.004
LSTAR(1,1)	0.004	0.014	0.052	0.086	0.009	0.022	0.028	0.060
LSTAR(2,1)	0.014	0.031	0.056	0.070	0.020	0.039	0.043	0.062
LSTAR(2,2)	**0.263**	**0.323**	**0.309**	**0.281**	**0.292**	**0.329**	**0.321**	**0.304**

(c) Overall Information Criteria

Table 4.5: Selection Frequencies of the Information Criteria: ESTAR vs. LSTAR models for LSTAR(2,2) DGP with $\gamma = 1$ and n=100

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,2)	0.000	0.000	0.148	0.153	0.000	0.000	0.139	0.146
ESTAR(2,3)	0.000	0.000	0.002	0.001	0.000	0.000	0.004	0.003
LSTAR(1,1)	0.035	0.049	0.000	0.000	0.037	0.053	0.000	0.000
LSTAR(2,1)	**0.285**	**0.333**	0.000	0.000	**0.288**	**0.333**	0.000	0.000
LSTAR(2,2)	0.201	0.246	**0.775**	**0.790**	0.205	0.250	**0.752**	**0.775**

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,2)	0.061	0.083	0.132	0.136	0.063	0.081	0.120	0.126
ESTAR(2,3)	0.009	0.006	0.004	0.002	0.009	0.006	0.006	0.004
LSTAR(1,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,2)	**0.601**	**0.740**	**0.836**	**0.850**	**0.611**	**0.746**	**0.820**	**0.845**

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,2)	0.059	0.074	0.089	0.092	0.060	0.075	0.087	0.089
ESTAR(2,3)	0.007	0.005	0.003	0.001	0.007	0.005	0.003	0.003
LSTAR(1,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,2)	**0.605**	**0.755**	**0.897**	**0.902**	**0.611**	**0.758**	**0.882**	**0.895**

(c) Overall Information Criteria

Table 4.6: Selection Frequencies of the Information Criteria:
ESTAR vs. LSTAR models for LSTAR(2,2) DGP with $\gamma = 1$ and n=1000

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.002	0.006	0.161	0.199	0.017	0.025	0.231	**0.284**
ESTAR(2,1)	0.005	0.006	0.012	0.017	0.009	0.010	0.025	0.029
ESTAR(2,2)	0.040	0.045	0.198	0.180	0.053	0.042	0.214	0.208
ESTAR(2,3)	0.005	0.003	0.008	0.004	0.002	0.002	0.009	0.002
LSTAR(1,1)	0.126	0.185	0.069	0.110	0.221	**0.283**	0.054	0.068
LSTAR(2,1)	0.128	0.158	0.051	0.064	0.175	0.206	0.033	0.045
LSTAR(2,2)	**0.239**	**0.276**	**0.306**	**0.289**	**0.241**	0.239	**0.250**	0.252

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.145	0.252	0.258	**0.348**	0.249	**0.318**	**0.308**	**0.363**
ESTAR(2,1)	0.005	0.009	0.008	0.009	0.019	0.029	0.013	0.019
ESTAR(2,2)	**0.274**	**0.321**	**0.331**	0.311	**0.256**	0.261	0.289	0.281
ESTAR(2,3)	0.044	0.028	0.017	0.007	0.016	0.008	0.014	0.002
LSTAR(1,1)	0.000	0.003	0.007	0.020	0.000	0.003	0.005	0.011
LSTAR(2,1)	0.003	0.005	0.012	0.016	0.009	0.017	0.012	0.018
LSTAR(2,2)	0.185	0.218	0.224	0.213	0.233	0.242	0.234	0.230

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.157	0.257	**0.374**	**0.439**	0.190	0.292	0.327	**0.385**
ESTAR(2,1)	0.005	0.008	0.011	0.013	0.007	0.008	0.008	0.012
ESTAR(2,2)	**0.304**	**0.349**	0.318	0.275	**0.355**	**0.351**	**0.339**	0.311
ESTAR(2,3)	0.051	0.032	0.008	0.003	0.045	0.023	0.021	0.007
LSTAR(1,1)	0.000	0.004	0.017	0.030	0.000	0.004	0.004	0.019
LSTAR(2,1)	0.002	0.004	0.014	0.017	0.002	0.009	0.012	0.017
LSTAR(2,2)	0.176	0.205	0.201	0.193	0.192	0.208	0.207	0.199

(c) Overall Information Criteria

Table 4.7: Selection Frequencies of the Information Criteria:
ESTAR vs. LSTAR models for LSTAR(2,2) DGP with $\gamma = 20$ and n=100

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.000	0.000	0.003	0.004	0.000	0.000	0.002	0.003
ESTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,2)	0.001	0.001	0.240	0.243	0.001	0.001	0.232	0.238
ESTAR(2,3)	0.000	0.000	0.005	0.003	0.000	0.000	0.007	0.005
LSTAR(1,1)	0.012	0.017	0.000	0.000	0.014	0.017	0.000	0.000
LSTAR(2,1)	0.133	0.164	0.000	0.000	0.133	0.165	0.000	0.000
LSTAR(2,2)	**0.254**	**0.315**	**0.694**	**0.710**	**0.261**	**0.318**	**0.670**	**0.695**

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.000	0.001	0.002	0.002	0.000	0.001	0.002	0.002
ESTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,2)	0.257	0.327	0.421	0.433	0.264	0.327	0.401	0.414
ESTAR(2,3)	0.045	0.037	0.013	0.008	0.049	0.037	0.016	0.013
LSTAR(1,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,2)	**0.389**	**0.480**	**0.538**	**0.545**	**0.401**	**0.489**	**0.540**	**0.549**

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
ESTAR(1,2)	0.000	0.001	0.003	0.004	0.000	0.001	0.003	0.003
ESTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ESTAR(2,2)	0.228	0.287	0.359	0.366	0.233	0.290	0.354	0.359
ESTAR(2,3)	0.042	0.032	0.012	0.007	0.040	0.032	0.013	0.012
LSTAR(1,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LSTAR(2,2)	**0.425**	**0.518**	**0.614**	**0.618**	**0.429**	**0.519**	**0.607**	**0.614**

(c) Overall Information Criteria

Table 4.8: Selection Frequencies of the Information Criteria:
ESTAR vs. LSTAR models for LSTAR(2,2) DGP with $\gamma = 20$ and n=1000

In the next step we also allow for SETAR models in the collection of models. As a result, we find another spurious behavior of the information criteria. In small samples all equally weighted and overall information criteria as well as the regime weighted SIC and WIC select the SETAR model. This might be due to the fact, that SETAR models have a smaller penalty term since the transition parameter γ does not have to be estimated. Hence, this spurious behavior can be interpreted as some kind of underfitting. The regime weighted versions of AIC and AICc prefer LSTAR models, even if the true DGP is a SETAR model. In larger samples the information criteria select the true model class. Again, the transition parameter γ is a key factor for the performance. For large values of γ the LSTAR model becomes a SETAR model. Then, it is impossible to distinguish these two model classes. The information criteria tend to favor the SETAR model in these cases, since the penalty term is smaller due to the missing transition parameter. For smaller values of γ the performance of the information criteria improves and especially for larger samples the differentiation between the different types of transition functions is reliable. Although in some cases a wrong model class is selected, the lag order coincides with the one preferred if only the correct model class is considered.
Finally, linear models are included into the collection of models. Then, the problem of distinguishing between linear and nonlinear becomes relevant again. As already shown, weighted versions of AIC and AICc cannot detect linearity. SIC and WIC tend to favor linear models although the model is nonlinear. This is especially a problem in small samples. The equally weighted and the overall versions of the information criteria often select the SETAR models. However, as already mentioned earlier, the performance improves with an increasing sample size. Regime weighted AIC and AICc spuriously select LSTAR models. Lag order selection works well. Even if not the correct model class is selected, the lag order combination corresponds to the artificial case with the correct model class.

Discriminating Regimes. Finally, we evaluate whether the information criteria can also be used to select the number of regimes. In this paragraph we only consider SETAR models with two or three regimes (cf. Gonzalo and Pitarakis, 2002; Clements and Krolzig, 1998) and reduce the maximum lag order to two, which yields eight lag order combinations for the three regime model. Furthermore, we also change the grid for thresholds. In this simulation study the grid consists of all quantiles from the 10% to the 90% quantile of y_t. Although this grid might be rough, we choose it because it guarantees that even if two consecutive grid points are chosen as thresholds, there will lie 10% of the observations in the middle regime. The estimation procedure of the thresholds is the sequential approach from Gonzalo and Pitarakis (2002). We consider the 1-step and the 2-step approach, i.e. we estimate the first threshold and given this threshold the second threshold. In the 1-step approach we keep these two estimates. In contrast, in the 2-step approach the first threshold is reestimated given the second threshold and finally the second threshold is reestimated given the refined first threshold (cf. Gonzalo and Pitarakis, 2002). For the three regime SETAR models the number of additional parameters increases to three (two thresholds, one delay parameter). Gonzalo and Pitarakis (2002) do not account for the number of thresholds in the penalty term, whereas Liu et al. (1997) incorporate thresholds as additional parameters into the penalty term.

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	**0.537**	**0.468**	0.000	0.000	**0.551**	**0.509**
SETAR(2,2)	0.000	0.000	0.040	0.031	0.004	0.004	0.048	0.040
SETAR(1,2,1)	0.134	0.176	0.034	0.040	**0.206**	**0.221**	0.024	0.029
SETAR(1,2,2)	0.190	**0.180**	0.033	0.034	0.173	0.160	0.029	0.029
SETAR(2,2,2)	**0.216**	0.162	0.035	0.029	0.113	0.088	0.022	0.018

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	**0.470**	**0.410**	0.000	0.000	**0.547**	**0.500**
SETAR(2,2)	0.000	0.000	0.043	0.027	0.001	0.001	0.057	0.043
SETAR(1,2,1)	0.104	0.165	0.051	0.057	**0.191**	**0.208**	0.018	0.020
SETAR(1,2,2)	0.194	**0.170**	0.031	0.030	0.179	0.157	0.014	0.014
SETAR(2,2,2)	**0.236**	0.146	0.034	0.019	0.100	0.074	0.017	0.012

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	**0.492**	**0.629**	**0.818**	**0.749**	**0.601**	**0.697**	**0.826**	**0.803**
SETAR(2,2)	0.072	0.045	0.001	0.000	0.065	0.036	0.007	0.003
SETAR(1,2,1)	0.042	0.027	0.000	0.000	0.034	0.017	0.001	0.001
SETAR(1,2,2)	0.045	0.018	0.000	0.000	0.026	0.011	0.002	0.001
SETAR(2,2,2)	0.071	0.015	0.000	0.000	0.031	0.004	0.000	0.000

(c) Overall Information Criteria

Table 4.9: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for AR(2) DGP with n=100;
1-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	**0.562**	**0.561**	0.000	0.000	**0.500**	**0.498**
SETAR(2,2)	0.000	0.000	0.085	0.085	0.000	0.000	0.089	0.083
SETAR(1,2,1)	0.009	0.009	0.019	0.020	0.009	0.009	0.020	0.021
SETAR(1,2,2)	0.119	0.126	0.055	0.055	0.119	0.127	0.053	0.059
SETAR(2,2,2)	**0.484**	**0.473**	0.119	0.115	**0.483**	**0.472**	0.144	0.137

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	**0.637**	**0.637**	0.000	0.000	**0.609**	**0.607**
SETAR(2,2)	0.000	0.000	0.036	0.030	0.000	0.000	0.042	0.037
SETAR(1,2,1)	0.002	0.004	0.011	0.018	0.002	0.005	0.007	0.011
SETAR(1,2,2)	0.034	0.059	0.053	0.051	0.035	0.060	0.043	0.057
SETAR(2,2,2)	**0.904**	**0.841**	0.158	0.137	**0.900**	**0.844**	0.220	0.187

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	**0.695**	**0.870**	**1.000**	**1.000**	**0.711**	**0.874**	**1.000**	**1.000**
SETAR(2,2)	0.146	0.083	0.000	0.000	0.145	0.081	0.000	0.000
SETAR(1,2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SETAR(1,2,2)	0.006	0.003	0.000	0.000	0.006	0.003	0.000	0.000
SETAR(2,2,2)	0.130	0.027	0.000	0.000	0.117	0.025	0.000	0.000

(c) Overall Information Criteria

Table 4.10: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for AR(2) DGP with n=1000;
1-step Estimation

Tables 4.9 and 4.10 present the selection frequencies of the information criteria if the true model is an AR(2) process and the 1-step approach is applied. Tables C.1 and C.2 show the respective results for the 2-step approach. Comparing these results, it becomes obvious that the weighted versions of the 1-step approach lead to more correct selections of the AR(2) model. The weighted versions of AIC and AICc never select the true model, but versions of SIC and WIC detect the linearity. The corresponding information criteria of the 2-step algorithm prefer a three regime SETAR model. Both approaches work well if the overall information criteria are applied. Then, the 2-step approach is even slightly superior.

In Tables 4.11, 4.12, C.3 and C.4 the results for the SETAR(1,1) process are tabulated.

In small samples weighted versions of SIC and WIC and the overall information criteria of the 1-step approach favor the linear models, whereas the respective versions of AIC and AICc favor the SETAR(1,1,1). For the 2-step approach equally weighted and overall SIC and WIC select the linear model. The other criteria prefer the SETAR(1,1,1). With an increasing sample size the weighted versions of SIC and WIC and all overall information criteria select the correct model when the 1-step approach is applied. For the 2-step approach only the overall criteria favor the correct model. The other versions prefer the SETAR(1,1,1).

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	**0.519**	**0.540**	0.000	0.000	**0.462**	**0.480**
SETAR(1,1)	0.005	0.006	0.177	0.200	0.012	0.015	0.222	0.252
SETAR(1,1,1)	**0.312**	**0.405**	0.089	0.096	**0.432**	**0.473**	0.074	0.079
SETAR(1,2,1)	0.177	0.178	0.030	0.026	0.188	0.173	0.038	0.031
SETAR(2,1,1)	0.172	0.161	0.030	0.026	0.144	0.135	0.021	0.019

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	**0.504**	**0.532**	0.000	0.000	**0.468**	**0.489**
SETAR(1,1)	0.000	0.000	0.117	0.145	0.003	0.004	0.245	0.272
SETAR(1,1,1)	**0.327**	**0.500**	0.119	0.159	**0.501**	**0.567**	0.062	0.064
SETAR(1,2,1)	0.179	0.157	0.064	0.041	0.168	0.150	0.053	0.050
SETAR(2,1,1)	0.160	0.142	0.040	0.033	0.129	0.121	0.016	0.018

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.147	**0.296**	**0.842**	**0.879**	**0.222**	**0.358**	**0.686**	**0.784**
SETAR(1,1)	0.167	0.288	0.096	0.090	0.217	0.315	0.171	0.147
SETAR(1,1,1)	**0.200**	0.158	0.012	0.005	0.174	0.117	0.030	0.018
SETAR(1,2,1)	0.094	0.046	0.001	0.000	0.070	0.029	0.007	0.002
SETAR(2,1,1)	0.088	0.036	0.002	0.001	0.057	0.024	0.007	0.004

(c) Overall Information Criteria

Table 4.11: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(1,1) DGP with n=100;
1-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.086	0.086	0.000	0.000	0.067	0.069
SETAR(1,1)	0.000	0.000	**0.472**	**0.479**	0.000	0.000	**0.441**	**0.455**
SETAR(1,1,1)	**0.419**	**0.496**	0.215	0.223	**0.436**	**0.504**	0.228	0.239
SETAR(1,2,1)	0.157	0.149	0.066	0.063	0.156	0.149	0.075	0.072
SETAR(2,1,1)	0.151	0.154	0.040	0.040	0.149	0.153	0.048	0.046

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.110	0.110	0.000	0.000	0.090	0.091
SETAR(1,1)	0.000	0.000	**0.524**	**0.536**	0.000	0.000	**0.506**	**0.524**
SETAR(1,1,1)	**0.469**	**0.634**	0.261	0.282	**0.504**	**0.650**	0.256	0.290
SETAR(1,2,1)	0.151	0.124	0.035	0.020	0.145	0.114	0.042	0.027
SETAR(2,1,1)	0.137	0.114	0.019	0.018	0.138	0.115	0.024	0.019

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.016	0.021	0.000	0.000	0.003	0.005
SETAR(1,1)	**0.306**	**0.535**	**0.953**	**0.967**	**0.321**	**0.546**	**0.940**	**0.963**
SETAR(1,1,1)	0.254	0.206	0.003	0.001	0.252	0.200	0.008	0.004
SETAR(1,2,1)	0.094	0.044	0.000	0.000	0.090	0.043	0.000	0.000
SETAR(2,1,1)	0.079	0.041	0.000	0.000	0.077	0.040	0.000	0.000

(c) Overall Information Criteria

Table 4.12: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(1,1) DGP with n=1000;
1-step Estimation

In Tables C.5 - C.8 in the Appendix the results for the SETAR(2,2,2) model can be found. Even in a small sample all versions of information criteria identify the correct model when the 1-step algorithm is applied. In large samples the selection frequencies converge towards 1. In case of the 2-step approach the weighted versions select the correct model class though not the correct lag order combination when the sample size is small. For a larger sample the correct model is selected. The overall information criteria favor a two regime SETAR model in small samples and an underfitted three regime SETAR model in the larger sample.

Taking into account all these results, we cannot find one superior approach. In general,

we would expect the 2-step algorithm to outperform the 1-step approach. However, this is not always case, which is probably due to the grid. Li (1988) points out that the choice of the grid has an effect on the threshold estimates. Thus, a finer grid might improve the results.
Generally, information criteria can be used to distinguish between SETAR models with a different number of regimes. The overall criteria perform quite well. But in small samples there is the possibility of underfitting. However, since the computational effort of estimating nonlinear models with more than two regimes is rather high, Gonzalo and Pitarakis (2002) recommend first to use a collection of models consisting only of linear and two regime models. If the information criteria select the nonlinear model, in a second step the two regime models and the three regime models are evaluated. If in the first step the linear model is preferred, it is not necessary to estimate the three regime models.

5 Conclusion

In this paper we evaluate the performance of different information criteria for simultaneous lag order and model class selection of nonlinear models. We focus on SETAR and STAR models due to the fact that they have a similar switching mechanism. Our set of information criteria consists of the commonly known criteria AIC, SIC and AICc. Furthermore, we also apply WIC which is supposed to perform well independent of the sample size. All in all, we consider 24 different information criteria with varying penalty terms, error term variance estimators and regime weightings. Our aim is to identify one or more criteria that can be used to select a best fitting model among different nonlinear model classes. Strictly speaking, information criteria cannot be employed in order to select between different model classes, because they are developed under the assumption that all models under consideration belong to the same parametric family (cf. Kapetanios, 2001). Nevertheless, this approach can be a valuable alternative to linearity tests. This is due to the fact that tests rely on a (possibly misspecified) lag order estimate, which influences the power properties of the tests (cf. Pitarakis, 2006). Furthermore, linearity tests with a specified alternative may have power against other models as well (cf. Luukkonen et al., 1988b). Hence, rejecting the null of a linear model does not tell which nonlinear model should be applied to the data. In contrast, model selection using information criteria will lead to a definite model choice.
Our results show that the information criteria perform well in general. However, there are some key factors that seriously influence the performance of the criteria. The *sample size* plays a crucial role. So, in small samples some criteria adopt a spurious behavior. Especially overall SIC and WIC tend to select a simple model. Depending on the

collection of models, this results in underfitting, the selection of linear models, although the true model is nonlinear, or a model with fewer additional parameters (cf. 4.3: Discriminating Nonlinear Models). In large samples the information criteria perform well. Another factor is the *identifiability* of the true model. One of its aspects concerns the lag order selection. If the largest lag has only a minor influence, the model is weakly identifiable (cf. McQuarrie et al., 1997) and underfitting occurs. Hence, underfitting is not necessarily a drawback of the information criteria but is due to the true data generating process. But also the *distance between regimes* affects the identifiability. If the regimes are very similar, a simpler model will be preferred. Especially in small samples the selection frequencies of the correct model are quite low. Asymptotically there are only few incorrect selections. The distance of regimes plays an important role for the discrimination between linear and nonlinear models. We have shown that weighted versions of AIC and AICc cannot detect linearity and therefore, should not be applied, if linear models are among the collection of models, since they spuriously point to the nonlinear model (cf. also Gonzalo and Pitarakis, 2002; Pitarakis, 2006). Another factor crucially influences the performance of the information criteria, as well. The *shape of the transition functions* of STAR models depends on the transition parameter γ. For small values both STAR models converge to AR models, whereas for large values the LSTAR model converges to a SETAR model and the ESTAR model reduces to a linear model. In these extreme cases it is difficult to distinguish the different model types. Due to the smaller penalties AR and SETAR models will be preferred to STAR models. Again a large sample size facilitates the selection process. For a larger number of observations, the differentiation of the type of transition function becomes more reliable (cf. also Psaradakis et al., 2009). The selection of the lag order combination does not suffer from misspecified model classes. Instead, it corresponds to the one selected if only the true model class is considered. Summarizing all our results, we cannot generally advise the use of one certain information criterion. Instead, several criteria should be applied in order to balance the individual strengths and weaknesses. However, our results show that the performance of the information criteria is not deteriorated if model specific parameters like threshold, delay or transition parameters are unknown and have to be estimated. Hence, the application of information criteria to empirical data in order to identify the best fitting model is an alternative approach to linearity tests. Though, weighted versions of AIC and AICc should not be used if linear and nonlinear model are compared. Moreover, the regime weighted criteria cannot outperform their equally weighted and overall counterparts. This is independent of the dominance of the regimes. Equally weighted and overall information criteria are more or less independent from the share of observations in each regime. The regime weighted criteria mostly perform better when the regimes are fairly separated. Hence, the equally weighted or overall versions should be preferred. The problem of choosing an adequate penalty term is similar to the application of information criteria to linear processes. Liu et al. (1997) remarks that for easily identifiable models a criterion with a large penalty should be applied in order to avoid overfitting, whereas for weakly identifiable models a criterion with a small penalty should be preferred to prevent underfitting. In fact, it is not clear whether the true model is weakly identifiable and thus, which information criterion should be applied. AIC and AICc are known to overfit. Our results show that the probability of overfitting can be reduced by using the unbiased error term variance estimator $\tilde{\sigma}^2$ (cf. also McQuarrie et al., 1997). Applying this estimator to SIC and WIC can lead to underfitting. Nevertheless, there are scenarios, where the selection frequencies of the correct model for SIC and WIC increase as well, when the unbiased estimator is used. Further modifications may lead to even better selection frequencies. So, Li (1988) points out that the choice of possible thresholds affects the final estimates. Therefore, a fine grid will probably lead to better threshold estimates and improve model selection. But in fact, the computational effort will increase enormously. This problem might be solved by applying a 2-step grid search algorithm. In the first step a rough grid is used to find a first threshold estimate. In the second step, a finer grid is built around this point and the estimate is refined. As a result, it would be unnecessary to do a global fine grid search. Instead, the computational effort would only increase locally.

A Data Generating Processes

AR(1)	$y_t = 0.5y_{t-1} + \varepsilon_t$
AR(2)	$y_t = -1.2 + 0.7y_{t-1} + 0.3y_{t-2} + \varepsilon_t$
AR(3)	$y_t = 1.2y_{t-1} - 0.35y_{t-2} - 0.1y_{t-3} + \varepsilon_t$

Table A.1: DGP: Autoregressive Processes

LSTAR(1,1)	$F_t(\cdot) = 1/(1 + \exp(-y_{t-1}))$
	$y_t = (0.8y_{t-1})(1 - F_t(\cdot)) + (0.2y_{t-1})F_t(\cdot) + \varepsilon_t$
LSTAR(2,2)	$F_t(\cdot) = 1/(1 + \exp(-y_{t-1}))$
	$y_t = (1.8y_{t-1} - 1.06y_{t-2})(1 - F_t(\cdot)) + (0.02 + 0.9y_{t-1} - 0.265y_{t-2})F_t(\cdot) + \varepsilon_t$
LSTAR(3,4)	$F_t(\cdot) = 1/(1 + \exp(-y_{t-1}))$
	$y_t = (1.8y_{t-1} - 1.06y_{t-2} - 0.2y_{t-3})(1 - F_t(\cdot))$
	$+(0.02 + 0.9y_{t-1} - 0.265y_{t-2} + 0.27y_{t-3} - 0.32y_{t-4})F_t(\cdot) + \varepsilon_t$

Table A.2: DGP: Logistic Smooth Transition Autoregressive Processes

ESTAR(1,1)	$F_t(\cdot) = 1 - \exp(-y_{t-1}^2)$
	$y_t = (0.8y_{t-1})(1 - F_t(\cdot)) + (0.2y_{t-1})F_t(\cdot) + \varepsilon_t$
ESTAR(2,2)	$F_t(\cdot) = 1 - \exp(-y_{t-1}^2)$
	$y_t = (1.8y_{t-1} - 1.06y_{t-2})(1 - F_t(\cdot)) + (0.02 + 0.9y_{t-1} - 0.265y_{t-2})F_t(\cdot) + \varepsilon_t$
ESTAR(3,4)	$F_t(\cdot) = 1 - \exp(-y_{t-1}^2)$
	$y_t = (1.8y_{t-1} - 1.06y_{t-2} - 0.2y_{t-3})(1 - F_t(\cdot))$
	$+(0.02 + 0.9y_{t-1} - 0.265y_{t-2} + 0.27y_{t-3} - 0.32y_{t-4})F_t(\cdot) + \varepsilon_t$

Table A.3: DGP: Exponential Smooth Transition Autoregressive Processes

SETAR(1,1)	$y_t = (0.8y_{t-1} + \varepsilon_t)\mathbb{I}_{y_{t-1}>0} + (0.2y_{t-1} + \varepsilon_t)\mathbb{I}_{y_{t-1}\leq 0}$
SETAR(1,3)	$y_t = (0.2y_{t-1}\varepsilon_t)\mathbb{I}_{y_{t-1}>0} + (1.2y_{t-1} - 0.35y_{t-2} - 0.1y_{t-3}\varepsilon_t)\mathbb{I}_{y_{t-1}\leq 0}$
SETAR(2,1)	$y_t = (1.0 + 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_t)\mathbb{I}_{y_{t-1}>0} + (0.8y_{t-1} + \varepsilon_t)\mathbb{I}_{y_{t-1}\leq 0}$
SETAR(2,2)	$y_t = (1.2 + 0.7y_{t-1} - 0.2y_{t-2} + \varepsilon_t)\mathbb{I}_{y_{t-1}>0}$
	$+ (1 - 1.1y_{t-1} + 0.18y_{t-2} + \varepsilon_t)\mathbb{I}_{y_{t-1}\leq 0}$
SETAR(3,2)	$y_t = (1.2y_{t-1} - 0.35y_{t-2} - 0.1y_{t-3} + \varepsilon_t)\mathbb{I}_{y_{t-1}>0}$
	$+ (1.2 + 0.7y_{t-1} - 0.2y_{t-2} + \varepsilon_t)\mathbb{I}_{y_{t-1}\leq 0}$
SETAR(4,3)	$y_t = (2.7607y_{t-1} - 3.8106y_{t-1} + 2.6535y_{t-3} - 0.9238y_{t-4} + \varepsilon_t)\mathbb{I}_{y_{t-1}>0}$
	$+ (1.2y_{t-1} - 0.35y_{t-2} - 0.1y_{t-3} + \varepsilon_t)\mathbb{I}_{y_{t-1}\leq 0}$
SETAR(1,1,1)	$y_t = (-0.5y_{t-1} + \varepsilon_t)\mathbb{I}_{y_{t-1}>0.5} + (0.8y_{t-1} + \varepsilon_t)\mathbb{I}_{-0.5<y_{t-1}\leq 0.5} + (0.2y_{t-1} + \varepsilon_t)\mathbb{I}_{y_{t-1}\leq -0.5}$
SETAR(2,2,2)	$y_t = (1 + 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_t)\mathbb{I}_{y_{t-2}>12} + (6 + 1.9y_{t-1} - 0.3y_{t-2} + \varepsilon_t)\mathbb{I}_{5<y_{t-2}\leq 12}$
	$+ (2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t)\mathbb{I}_{y_{t-2}\leq 5}$

Table A.4: DGP: Self-exciting Threshold Autoregressive Processes

B Discriminating Linear and Nonlinear Model

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.371	0.379	0.000	0.000	0.286	0.306
SETAR(1,1)	**0.295**	**0.400**	**0.390**	**0.423**	**0.322**	**0.420**	**0.391**	**0.440**
SETAR(1,2)	0.096	0.107	0.075	0.068	0.102	0.101	0.089	0.081
SETAR(1,3)	0.084	0.067	0.034	0.025	0.081	0.064	0.043	0.034
SETAR(1,4)	0.075	0.065	0.025	0.020	0.069	0.060	0.035	0.024
SETAR(2,1)	0.110	0.114	0.051	0.049	0.112	0.112	0.059	0.053
SETAR(3,1)	0.071	0.068	0.021	0.017	0.070	0.068	0.033	0.026
SETAR(4,1)	0.067	0.053	0.012	0.008	0.063	0.052	0.016	0.011

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.406	0.411	0.000	0.000	0.356	0.366
SETAR(1,1)	**0.371**	**0.634**	**0.478**	**0.527**	**0.448**	**0.679**	**0.471**	**0.540**
SETAR(1,2)	0.106	0.113	0.041	0.026	0.114	0.094	0.049	0.034
SETAR(1,3)	0.074	0.047	0.011	0.002	0.064	0.039	0.018	0.004
SETAR(1,4)	0.081	0.024	0.007	0.001	0.054	0.020	0.012	0.003
SETAR(2,1)	0.114	0.092	0.034	0.022	0.115	0.090	0.049	0.031
SETAR(3,1)	0.070	0.043	0.009	0.003	0.063	0.039	0.016	0.008
SETAR(4,1)	0.050	0.018	0.005	0.003	0.040	0.016	0.008	0.004

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.011	0.042	**0.632**	**0.691**	0.012	0.050	0.442	**0.512**
SETAR(1,1)	**0.388**	**0.641**	0.327	0.286	**0.425**	**0.645**	**0.471**	0.440
SETAR(1,2)	0.108	0.094	0.011	0.004	0.109	0.090	0.027	0.014
SETAR(1,3)	0.071	0.037	0.000	0.000	0.068	0.037	0.003	0.000
SETAR(1,4)	0.078	0.025	0.001	0.000	0.066	0.022	0.002	0.001
SETAR(2,1)	0.112	0.085	0.014	0.011	0.114	0.085	0.029	0.018
SETAR(3,1)	0.068	0.033	0.002	0.000	0.062	0.033	0.006	0.002
SETAR(4,1)	0.048	0.018	0.001	0.001	0.045	0.017	0.001	0.001

(c) Overall Information Criteria

Table B.1: Selection Frequencies of the Information Criteria:
AR vs. SETAR models for SETAR(1,1) DGP with n=250

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.203	0.210	0.000	0.000	0.173	0.181
SETAR(1,1)	**0.349**	**0.447**	**0.540**	**0.567**	**0.367**	**0.449**	**0.514**	**0.554**
SETAR(1,2)	0.128	0.118	0.096	0.086	0.125	0.118	0.109	0.096
SETAR(1,3)	0.079	0.068	0.042	0.037	0.078	0.067	0.054	0.046
SETAR(1,4)	0.082	0.068	0.027	0.025	0.078	0.068	0.032	0.027
SETAR(2,1)	0.088	0.088	0.049	0.042	0.090	0.089	0.058	0.053
SETAR(3,1)	0.068	0.058	0.020	0.018	0.065	0.058	0.025	0.020
SETAR(4,1)	0.057	0.046	0.005	0.003	0.054	0.046	0.006	0.005

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.244	0.246	0.000	0.000	0.220	0.226
SETAR(1,1)	**0.444**	**0.692**	**0.691**	**0.716**	**0.489**	**0.712**	**0.662**	**0.712**
SETAR(1,2)	0.118	0.109	0.035	0.025	0.118	0.109	0.054	0.033
SETAR(1,3)	0.072	0.031	0.004	0.000	0.068	0.028	0.008	0.003
SETAR(1,4)	0.067	0.025	0.001	0.001	0.052	0.021	0.008	0.001
SETAR(2,1)	0.089	0.073	0.017	0.010	0.086	0.068	0.029	0.018
SETAR(3,1)	0.066	0.030	0.002	0.000	0.059	0.026	0.006	0.002
SETAR(4,1)	0.047	0.018	0.002	0.000	0.041	0.017	0.004	0.002

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.001	0.003	0.279	0.317	0.001	0.003	0.144	0.190
SETAR(1,1)	**0.480**	**0.714**	**0.687**	**0.661**	**0.500**	**0.723**	**0.793**	**0.777**
SETAR(1,2)	0.122	0.101	0.022	0.015	0.123	0.101	0.036	0.023
SETAR(1,3)	0.060	0.032	0.000	0.000	0.058	0.031	0.003	0.000
SETAR(1,4)	0.067	0.023	0.000	0.000	0.064	0.019	0.000	0.000
SETAR(2,1)	0.076	0.063	0.007	0.003	0.077	0.061	0.018	0.009
SETAR(3,1)	0.059	0.028	0.000	0.000	0.056	0.026	0.003	0.000
SETAR(4,1)	0.044	0.019	0.000	0.000	0.041	0.019	0.000	0.000

(c) Overall Information Criteria

Table B.2: Selection Frequencies of the Information Criteria: AR vs. SETAR models for SETAR(1,1) DGP with n=500

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.030	0.031	0.000	0.000	0.016	0.021
AR(2)	0.000	0.000	0.042	0.041	0.000	0.000	0.041	0.038
SETAR(1,1)	0.172	**0.206**	0.131	0.173	0.184	**0.215**	0.097	0.136
SETAR(1,2)	0.161	0.195	0.186	0.219	0.172	0.204	0.162	0.192
SETAR(1,3)	0.071	0.064	0.035	0.033	0.065	0.059	0.034	0.034
SETAR(2,1)	0.081	0.101	0.134	0.138	0.083	0.102	0.133	0.154
SETAR(2,2)	**0.173**	0.182	**0.285**	**0.252**	**0.185**	0.188	**0.310**	**0.283**
SETAR(2,3)	0.059	0.050	0.070	0.051	0.053	0.045	0.084	0.062

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.040	0.043	0.000	0.000	0.021	0.032
AR(2)	0.000	0.000	0.062	0.056	0.000	0.000	0.063	0.056
SETAR(1,1)	0.019	0.053	0.103	0.165	0.029	0.063	0.069	0.127
SETAR(1,2)	0.070	0.126	0.180	0.213	0.086	0.146	0.157	0.191
SETAR(1,3)	0.027	0.026	0.023	0.017	0.023	0.024	0.022	0.021
SETAR(2,1)	0.056	0.115	0.109	0.118	0.082	0.126	0.106	0.123
SETAR(2,2)	**0.337**	**0.422**	**0.363**	**0.327**	**0.388**	**0.432**	**0.393**	**0.356**
SETAR(2,3)	0.142	0.095	0.061	0.035	0.110	0.078	0.075	0.049

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.070	0.095	0.000	0.000	0.024	0.044
AR(2)	0.002	0.007	0.128	0.137	0.002	0.010	0.084	0.095
SETAR(1,1)	0.020	0.051	0.119	0.149	0.021	0.053	0.099	0.142
SETAR(1,2)	0.079	0.129	0.190	0.208	0.081	0.135	0.175	0.197
SETAR(1,3)	0.027	0.030	0.014	0.013	0.029	0.028	0.014	0.016
SETAR(2,1)	0.047	0.104	0.115	0.112	0.057	0.105	0.130	0.133
SETAR(2,2)	**0.346**	**0.433**	**0.315**	**0.256**	**0.368**	**0.436**	**0.378**	**0.323**
SETAR(2,3)	0.143	0.099	0.026	0.015	0.143	0.097	0.051	0.026

(c) Overall Information Criteria

Table B.3: Selection Frequencies of the Information Criteria:
AR vs. SETAR models for SETAR(2,3) DGP with n=250

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.002	0.002	0.000	0.000	0.001	0.001
AR(2)	0.000	0.000	0.004	0.004	0.000	0.000	0.003	0.003
SETAR(1,1)	**0.210**	**0.237**	0.039	0.057	**0.216**	0.238	0.029	0.040
SETAR(1,2)	0.181	0.201	0.100	0.114	0.187	0.203	0.071	0.098
SETAR(1,3)	0.060	0.046	0.010	0.009	0.057	0.047	0.013	0.010
SETAR(2,1)	0.061	0.072	0.169	0.190	0.062	0.074	0.156	0.173
SETAR(2,2)	0.185	0.235	**0.514**	**0.504**	0.197	**0.242**	**0.522**	**0.522**
SETAR(2,3)	0.070	0.056	0.103	0.071	0.067	0.052	0.121	0.094

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.003	0.003	0.000	0.000	0.001	0.001
AR(2)	0.000	0.000	0.014	0.013	0.000	0.000	0.012	0.013
SETAR(1,1)	0.002	0.006	0.019	0.027	0.002	0.007	0.013	0.021
SETAR(1,2)	0.010	0.025	0.066	0.081	0.014	0.027	0.045	0.071
SETAR(1,3)	0.001	0.002	0.002	0.004	0.001	0.003	0.004	0.002
SETAR(2,1)	0.030	0.050	0.101	0.153	0.032	0.058	0.074	0.116
SETAR(2,2)	**0.465**	**0.599**	**0.672**	**0.645**	**0.501**	**0.618**	**0.675**	**0.671**
SETAR(2,3)	0.182	0.148	0.092	0.058	0.167	0.131	0.117	0.078

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.001
AR(2)	0.001	0.001	0.014	0.022	0.001	0.001	0.006	0.010
SETAR(1,1)	0.002	0.007	0.028	0.041	0.002	0.007	0.022	0.028
SETAR(1,2)	0.010	0.021	0.089	0.105	0.011	0.023	0.058	0.091
SETAR(1,3)	0.001	0.003	0.002	0.004	0.001	0.003	0.003	0.002
SETAR(2,1)	0.023	0.046	0.130	0.155	0.024	0.046	0.095	0.131
SETAR(2,2)	**0.468**	**0.606**	**0.677**	**0.639**	**0.481**	**0.616**	**0.711**	**0.677**
SETAR(2,3)	0.192	0.144	0.044	0.025	0.192	0.140	0.072	0.045

(c) Overall Information Criteria

Table B.4: Selection Frequencies of the Information Criteria:
AR vs. SETAR models for SETAR(2,3) DGP with n=500

C Discriminating Regimes

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	0.115	0.097	0.000	0.000	0.136	0.125
SETAR(2,2)	0.000	0.000	0.008	0.006	0.000	0.000	0.010	0.007
SETAR(1,2,1)	0.136	0.156	0.109	0.115	0.155	0.167	0.100	0.110
SETAR(1,2,2)	**0.213**	**0.207**	**0.168**	**0.167**	**0.209**	**0.202**	**0.177**	**0.173**
SETAR(2,2,2)	0.152	0.121	0.105	0.089	0.108	0.100	0.091	0.082

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	**0.157**	0.139	0.000	0.000	**0.235**	**0.220**
SETAR(2,2)	0.000	0.000	0.010	0.006	0.000	0.000	0.014	0.010
SETAR(1,2,1)	0.103	0.114	0.084	0.092	0.136	0.138	0.062	0.066
SETAR(1,2,2)	**0.189**	**0.191**	0.154	**0.152**	**0.181**	**0.173**	0.133	0.129
SETAR(2,2,2)	0.167	0.129	0.097	0.089	0.101	0.086	0.074	0.061

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	**0.536**	**0.645**	**0.818**	**0.749**	**0.622**	**0.703**	**0.828**	**0.804**
SETAR(2,2)	0.120	0.058	0.001	0.000	0.090	0.041	0.007	0.003
SETAR(1,2,1)	0.020	0.007	0.000	0.000	0.015	0.002	0.000	0.000
SETAR(1,2,2)	0.016	0.007	0.000	0.000	0.010	0.005	0.000	0.000
SETAR(2,2,2)	0.024	0.005	0.000	0.000	0.011	0.001	0.000	0.000

(c) Overall Information Criteria

Table C.1: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for AR(2) DGP with n=100;
2-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	0.059	0.059	0.000	0.000	0.053	0.052
SETAR(2,2)	0.000	0.000	0.005	0.005	0.000	0.000	0.007	0.006
SETAR(1,2,1)	0.011	0.012	0.003	0.003	0.011	0.012	0.003	0.003
SETAR(1,2,2)	**0.389**	**0.390**	**0.529**	**0.534**	**0.390**	**0.390**	**0.526**	**0.529**
SETAR(2,2,2)	0.302	0.301	0.194	0.189	0.301	0.301	0.198	0.195

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	0.000	0.000	0.116	0.116	0.000	0.000	0.106	0.105
SETAR(2,2)	0.000	0.000	0.001	0.001	0.000	0.000	0.006	0.004
SETAR(1,2,1)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
SETAR(1,2,2)	**0.443**	**0.446**	**0.426**	**0.431**	**0.443**	**0.445**	**0.423**	**0.429**
SETAR(2,2,2)	0.347	0.341	0.259	0.251	0.347	0.340	0.269	0.260

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(2)	**0.734**	**0.880**	**1.000**	**1.000**	**0.744**	**0.884**	**1.000**	**1.000**
SETAR(2,2)	0.208	0.094	0.000	0.000	0.203	0.091	0.000	0.000
SETAR(1,2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SETAR(1,2,2)	0.002	0.000	0.000	0.000	0.002	0.000	0.000	0.000
SETAR(2,2,2)	0.032	0.006	0.000	0.000	0.027	0.006	0.000	0.000

(c) Overall Information Criteria

Table C.2: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·,·) vs. SETAR(3;·,·,·) models for AR(2) DGP with n=1000;
2-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.154	0.164	0.000	0.000	0.144	0.154
SETAR(1,1)	0.015	0.017	0.061	0.067	0.023	0.026	0.086	0.091
SETAR(1,1,1)	**0.205**	**0.252**	**0.161**	**0.177**	**0.254**	**0.289**	**0.158**	**0.167**
SETAR(1,2,1)	0.182	0.189	0.141	0.142	0.182	0.176	0.150	0.141
SETAR(2,1,1)	0.175	0.176	0.146	0.145	0.173	0.170	0.150	0.151

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	**0.200**	**0.215**	0.000	0.000	**0.220**	**0.234**
SETAR(1,1)	0.000	0.000	0.046	0.057	0.003	0.003	0.095	0.101
SETAR(1,1,1)	**0.219**	**0.283**	0.148	0.170	**0.288**	**0.327**	0.123	0.133
SETAR(1,2,1)	0.177	0.177	0.145	0.141	0.195	0.189	0.155	0.146
SETAR(2,1,1)	0.186	0.189	0.145	0.141	0.177	0.170	0.129	0.129

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.179	0.310	**0.845**	**0.880**	0.237	**0.380**	**0.698**	**0.789**
SETAR(1,1)	**0.286**	**0.379**	0.103	0.092	**0.319**	0.376	0.188	0.158
SETAR(1,1,1)	0.091	0.069	0.005	0.003	0.076	0.050	0.010	0.006
SETAR(1,2,1)	0.054	0.026	0.000	0.000	0.034	0.021	0.005	0.001
SETAR(2,1,1)	0.038	0.015	0.000	0.000	0.024	0.007	0.001	0.001

(c) Overall Information Criteria

Table C.3: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(1,1) DGP with n=100;
2-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.018	0.018	0.000	0.000	0.015	0.016
SETAR(1,1)	0.034	0.036	0.162	0.164	0.034	0.037	0.153	0.157
SETAR(1,1,1)	**0.296**	**0.341**	**0.276**	**0.287**	**0.309**	**0.343**	**0.273**	**0.283**
SETAR(1,2,1)	0.159	0.149	0.136	0.136	0.156	0.149	0.139	0.138
SETAR(2,1,1)	0.215	0.210	0.187	0.184	0.214	0.212	0.191	0.187

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.030	0.030	0.000	0.000	0.027	0.028
SETAR(1,1)	0.000	0.000	0.115	0.116	0.000	0.000	0.107	0.113
SETAR(1,1,1)	**0.328**	**0.381**	**0.318**	**0.330**	**0.342**	**0.383**	**0.311**	**0.323**
SETAR(1,2,1)	0.175	0.164	0.146	0.146	0.173	0.163	0.147	0.146
SETAR(2,1,1)	0.226	0.216	0.193	0.190	0.223	0.219	0.196	0.193

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
AR(1)	0.000	0.000	0.016	0.021	0.000	0.000	0.003	0.005
SETAR(1,1)	**0.508**	**0.692**	**0.955**	**0.967**	**0.515**	**0.698**	**0.942**	**0.965**
SETAR(1,1,1)	0.109	0.083	0.001	0.001	0.109	0.081	0.006	0.002
SETAR(1,2,1)	0.046	0.021	0.000	0.000	0.045	0.020	0.001	0.000
SETAR(2,1,1)	0.023	0.010	0.000	0.000	0.022	0.010	0.000	0.000

(c) Overall Information Criteria

Table C.4: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(1,1) DGP with n=1000;
2-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.060	0.087	0.084	0.110	0.129	0.149	0.133	0.147
SETAR(2,1,2)	0.139	0.130	0.128	0.121	0.134	0.133	0.130	0.128
SETAR(2,2,1)	0.189	0.226	0.259	0.306	0.166	0.203	0.190	0.241
SETAR(2,2,2)	**0.554**	**0.492**	**0.467**	**0.397**	**0.484**	**0.432**	**0.459**	**0.398**

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.050	0.080	0.081	0.106	0.121	0.141	0.129	0.143
SETAR(2,1,2)	0.117	0.106	0.101	0.097	0.111	0.110	0.107	0.104
SETAR(2,2,1)	0.201	0.240	0.277	0.321	0.179	0.216	0.205	0.258
SETAR(2,2,2)	**0.552**	**0.488**	**0.461**	**0.392**	**0.481**	**0.425**	**0.451**	**0.388**

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.019	0.035	0.052	0.067	0.026	0.040	0.043	0.056
SETAR(2,1,2)	0.017	0.013	0.012	0.010	0.019	0.013	0.013	0.012
SETAR(2,2,1)	0.344	**0.452**	**0.501**	**0.542**	**0.402**	**0.472**	**0.480**	**0.514**
SETAR(2,2,2)	**0.434**	0.308	0.236	0.178	0.365	0.279	0.268	0.217

(c) Overall Information Criteria

Table C.5: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(2,2,2) DGP with n=100;
1-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.001	0.001	0.005	0.005	0.001	0.001	0.004	0.005
SETAR(2,1,2)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
SETAR(2,2,1)	0.010	0.012	0.018	0.020	0.010	0.012	0.016	0.018
SETAR(2,2,2)	**0.978**	**0.976**	**0.968**	**0.966**	**0.978**	**0.976**	**0.970**	**0.968**

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
SETAR(2,1,2)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
SETAR(2,2,1)	0.015	0.017	0.023	0.025	0.015	0.017	0.021	0.023
SETAR(2,2,2)	**0.969**	**0.967**	**0.961**	**0.959**	**0.969**	**0.967**	**0.963**	**0.961**

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,2,1)	0.017	0.028	0.059	0.076	0.017	0.028	0.050	0.057
SETAR(2,2,2)	**0.979**	**0.967**	**0.936**	**0.919**	**0.979**	**0.967**	**0.945**	**0.938**

(c) Overall Information Criteria

Table C.6: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(2,2,2) DGP with n=1000;
1-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.102	0.142	0.124	0.183	0.198	0.224	0.186	0.217
SETAR(2,1,2)	**0.376**	**0.415**	**0.323**	**0.358**	**0.397**	**0.396**	**0.325**	**0.332**
SETAR(2,2,1)	0.169	0.165	0.206	0.181	0.111	0.111	0.143	0.152
SETAR(2,2,2)	0.285	0.206	0.261	0.192	0.225	0.200	0.266	0.222

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.106	0.149	0.132	0.190	0.205	0.242	0.191	0.227
SETAR(2,1,2)	**0.371**	**0.410**	**0.325**	**0.361**	**0.396**	**0.394**	**0.320**	**0.328**
SETAR(2,2,1)	0.164	0.159	0.199	0.174	0.106	0.101	0.139	0.146
SETAR(2,2,2)	0.282	0.208	0.257	0.190	0.203	0.183	0.246	0.208

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1)	0.147	0.200	0.263	**0.300**	0.163	0.217	0.244	0.270
SETAR(2,2)	**0.402**	**0.347**	**0.295**	0.255	**0.386**	**0.333**	**0.309**	**0.283**
SETAR(2,1,1)	0.082	0.089	0.089	0.089	0.082	0.088	0.089	0.090
SETAR(2,1,2)	0.012	0.006	0.004	0.004	0.011	0.007	0.006	0.004
SETAR(2,2,1)	0.095	0.104	0.105	0.103	0.102	0.104	0.104	0.105
SETAR(2,2,2)	0.069	0.045	0.022	0.015	0.054	0.038	0.030	0.019

(c) Overall Information Criteria

Table C.7: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(2,2,2) DGP with n=100;
2-step Estimation

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
SETAR(2,1,2)	0.019	0.038	0.050	0.065	0.027	0.044	0.044	0.053
SETAR(2,2,1)	0.021	0.023	0.025	0.026	0.021	0.023	0.025	0.025
SETAR(2,2,2)	**0.954**	**0.933**	**0.919**	**0.903**	**0.946**	**0.927**	**0.925**	**0.916**

(a) Regime Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,1,1)	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
SETAR(2,1,2)	0.019	0.038	0.050	0.065	0.027	0.044	0.044	0.053
SETAR(2,2,1)	0.020	0.022	0.024	0.025	0.020	0.022	0.024	0.024
SETAR(2,2,2)	**0.952**	**0.931**	**0.917**	**0.901**	**0.944**	**0.925**	**0.923**	**0.914**

(b) Equally Weighted Information Criteria

	$AIC_{\hat{\sigma}^2}$	$AIC_{\tilde{\sigma}^2}$	$SIC_{\hat{\sigma}^2}$	$SIC_{\tilde{\sigma}^2}$	$AICc_{\hat{\sigma}^2}$	$AICc_{\tilde{\sigma}^2}$	$WIC_{\hat{\sigma}^2}$	$WIC_{\tilde{\sigma}^2}$
SETAR(2,2)	0.449	0.449	0.449	0.449	0.449	0.449	0.449	0.449
SETAR(2,1,1)	**0.482**	**0.500**	**0.504**	**0.504**	**0.483**	**0.501**	**0.504**	**0.504**
SETAR(2,1,2)	0.021	0.003	0.000	0.000	0.020	0.002	0.000	0.000
SETAR(2,2,1)	0.015	0.018	0.019	0.019	0.015	0.018	0.019	0.019
SETAR(2,2,2)	0.016	0.013	0.011	0.011	0.016	0.013	0.011	0.011

(c) Overall Information Criteria

Table C.8: Selection Frequencies of the Information Criteria:
AR vs. SETAR(2;·,·) vs. SETAR(3;·,·,·) models for SETAR(2,2,2) DGP with n=1000;
2-step Estimation

References

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

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

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.

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

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.

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

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.

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

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*, 75:491–499.

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

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.

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.

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)*, pages 245–292.

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.