Rossen, Anja

Working Paper

On the predictive content of nonlinear transformations of lagged autoregression residuals and time series observations

HWWI research paper, No. 113

Provided in cooperation with:
Hamburgisches WeltWirtschaftsInstitut (HWWI)

Hamburg Institute of International Economics

On the predictive content of nonlinear transformations of lagged autoregression residuals and time series observations

Anja Rossen

HWWI Research

Paper 113

Hamburg Institute of International Economics (HWWI) | 2011
ISSN 1861-504X

Anja Rossen
Hamburg Institute of International Economics (HWWI)
Heimhuder Str. 71 | 20148 Hamburg | Germany
Phone: +49 (0)40 34 05 76 - 347 | Fax: +49 (0)40 34 05 76 - 776
rossen@hwwi.org

HWWI Research Paper
Hamburg Institute of International Economics (HWWI)
Heimhuder Str. 71 | 20148 Hamburg | Germany
Phone: +49 (0)40 34 05 76 - 0 | Fax: +49 (0)40 34 05 76 - 776
info@hwwi.org | www.hwwi.org
ISSN 1861-504X

On the predictive content of nonlinear transformations of lagged autoregression residuals and time series observations

Anja Rossen[†]

Hamburg Institute of International Economics (HWWI)
October 2011

Abstract

This study focuses on the question whether nonlinear transformation of lagged time series values and residuals are able to systematically improve the average forecasting performance of simple Autoregressive models. Furthermore it investigates the potential superior forecasting results of a nonlinear Threshold model. For this reason, a large-scale comparison over almost 400 time series which span from 1996:3 up to 2008:12 (production indices, price indices, unemployment rates, exchange rates, money supply) from 10 European countries is made. The average forecasting performance is appraised by means of Mean Group statistics and simple t-tests. Autoregressive models are extended by transformed first lags of residuals and time series values. Whereas additional transformation of lagged time series values are able to reduce the ex-ante forecast uncertainty and provide a better directional accuracy, transformations of lagged residuals also lead to smaller forecast errors. Furthermore, the nonlinear Threshold model is able to capture certain type of economic behavior in the data and provides superior forecasting results than a simple Autoregressive model. These findings are widely independent of considered economic variables.

Keywords: Time series modeling, forecasting comparison, nonlinear transformations, Threshold Autoregressive modeling, average forecasting performance

JEL Classification: C22, C53, C51

[†]Hamburg Institute of International Economics (HWWI), Heimhuder Str. 71, 20148 Hamburg
E-mail: rossen@hwwi.org, Tel. +49 (0)40 34 05 76 - 347, Fax +49 (0)40 34 05 76 - 776

1 Introduction

Forecasting is a major focus in empirical economics. A researcher making a time series forecast is confronted with a quantity of possible models, estimations procedures and forecasting methods. These questions thus arise: Which model provides an optimal approximation for a considered time series of interest and which forecasting method is a-priori a good choice with respect to its forecasting performance?

Linear models are widely used and supply good forecasting results. But still, one could think that these models are not able to capture certain types of economic behavior in the data. Nonlinear models have become more common in recent years and an increased interest in forecasting economic variables with nonlinear models has arisen. Large-scale comparisons of the forecasting performance of linear and nonlinear models have been appeared in the literature (for example, see Marcellino et al., 2006 and Teräsvrita et al., 2003). There is no clear agreement whether nonlinear or linear models perform better concerning the out-of-sample forecasting results. A Monte Carlo study by Clements and Smith (1999) comes to the result that nonlinear models not always outperform linear models but are favorably when the forecast origin happens to be in a certain state of the the process. Nonlinear features that are presented in the data may not persists in the future and a good in-sample fit does not necessarily induce a good out-of-sample forecasting performance (Diebold and Nason, 1990).

It is obvious, that nonlinear models give an important contribution to forecasting economic variables. Another field of research, forecasting transformed time series, has also a great interest in the literature. By means of nonlinear transformations a forecaster attempts to obtain a time series with 'better' properties in order to get improved forecasting results. Such a transformation, like the logarithm, may inherent informations that are improving the forecasting performance of the level of an economic variable as well. The often employed logarithm function is beneficial for forecasting if it is leading to a more Gaussian process. But, converting an optimal forecast of the logarithm back to forecasts for the original variable (via the exponential function), is not always suitable (Lütkepohl and Xu, 2009). If an optimal forecast for a transformed time series exists, it should be used (Granger and Newbold, 1976).

This study combines both forecasting issues to a new direction of research. It investigates whether and under which circumstances a certain nonlinear transformation of lagged time series values or lagged residuals can a-priori help to systematically improve the forecasting performance of a simple linear Autoregressive model. The goal is to find a certain transformed Autoregressive model that most frequently leads to superior forecasting results. Such a transformed model may perform the best for certain types of economic variables. Furthermore, this study examines and compares these results to the forecasting performance of a simple nonlinear Threshold model. Therefore, a large-scale empirical comparison of forecasting models with various nonlinear transformations, using data on 382 monthly time series of 10 European economies is made. Instead of focusing on single variables, the average forecasting performance over all time series and economies is considered. Using this data, models with data-dependent lag order selection like the AIC and the BIC are used. Expanding and rolling estimation window are applied and one-step ahead forecasts are recursively iterated forward for 23 forecasting steps. To make a stable statement on the predictive content of transformed Autoregressive models five different loss functions, Mean Group statistics (MG) and inference are evaluated. Furthermore, all models are investigated whether their results can be carried over to different subsamples of time series.

The remainder of this paper is organized as follows. In Section two, three Autoregressive models are presented. Their estimation procedure and model selection techniques are briefly discussed. Section three gives an extensively description of the data set. The empirical application and the forecasting comparison are documented in section four. The next section discusses the empirical results for full- and subsample evaluations. Section six contains a conclusion.

2 Three Autoregressive models and its model selection procedures

2.1 Simple Autoregressive models

Let y_t denote a stationary time series of interest. An univariate Autoregressive process of order p (AR(p)) is given by:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + e_t, \quad t = 1, ..., N, \tag{2.1}$$

where α denotes an intercept, p the lag order and $e_t \sim iid(0, \sigma_e^2)$ is a homoscedastic white noise process with zero mean and variance σ_e^2. For a given lag order p, parameters $\alpha, \beta_1, .., \beta_p$ and σ_e^2 are estimated by Ordinary Least Squares. Nevertheless, lag order p is usually unknown and will be estimated by means of two simple and commonly used information criterions (IC):

$$IC(p) = \underbrace{log(\hat{\sigma}_e^2)}_{\substack{goodness \\ of fit}} + \underbrace{\frac{c^*(p+1)}{N}}_{\substack{penalty \\ term}}, \tag{2.2}$$

where c^*= 2 for the Aikake Information Criterion (AIC) or c^*=$log(N)$ for the Bayesian Information Criterion (BIC), respectively. $\hat{\sigma}_e^2$ is the estimated residual variance for a particular lag order choice

$p \in (1, ..., p_{max})$ and N is the sample size. Whereas, Monte Carlo studies (Jones, 1975 and Ohtani, 2003) show, that the AIC criterion has a tendency to overestimate lag order p and leads to complex and over-fitted models the Bayesian information criterion considers the issue of over-fitting and includes a stronger penalty term.

The second considered model in this study is an Autoregressive Moving Average process of orders p and q (ARMA(p,q)). It‘s general representation is given by:

$$\tilde{y}_t = \sum_{i=1}^{p} \beta_i \tilde{y}_{t-i} + \sum_{j=1}^{q} \delta_j e_{t-j} + e_t, \quad t = 1, ..., N. \tag{2.3}$$

Again, $e_t \sim iid(0, \sigma_e^2)$ a homoscedastic white noise process with zero mean and variance σ_e^2. $\tilde{y}_t = y_t - \bar{y}$ is a stationary, mean adjusted time series of interest. For simplification, a common technique is used and unobservable residuals in the first estimation step are equally set to their mean zero: $e_t = 0$ for $t \leq 0$ and $e_t = y_t - \sum_{i=1}^{p} \beta_i y_{t-i} - \sum_{j=1}^{q} \delta_j e_{t-j}$ for $t > 0$ (for example, see Schlittgen & Streitberg, 2001).

Again, lag orders $p \in (1, ..., p_{max})$ and $q \in (1, .., q_{max})$ are unknown and selected by means of both information criterions, simply by replacing the common penalty term $\frac{c^*(p+1)}{N}$ by $\frac{c^*(p+q)}{N}$.

2.2 (Self-exciting) Threshold Autoregressive model

So far, simple linear Autoregressive models have been introduced. Although, these models tend to make a good job in fitting and forecasting data, they are still an approximation and are not always able to present certain features in the data. In contrast to this, nonlinear models are usually able to capture features like asymmetry, limit cycles or amplitude-frequency dependency. A simple and quite popular nonlinear model, the Threshold Autoregressive model (TAR), was first introduced by Tong and Lim (1980)[1]. This model is based on the idea of a piecewise linearization over the state space. Depending on a so called threshold variable relative to a threshold value, coefficients of a linear Autoregressive process and hence the linear relationship can vary across different regimes. Accordingly, a Threshold Autoregressive model is locally linear in the threshold space.

A special case of the Threshold model appears if the threshold variable is defined as past values of the time series itself. The resulting model is called a (Self-Exciting) Threshold Autoregressive model (SETAR) and is given by:

$$y_t = \alpha^j + \beta_1^j y_{t-1} + \beta_2^j y_{t-2} + ... + \beta_{p_j}^j y_{t-p_j} + e_t^j \qquad \text{if} \quad q_{j-1} \leq y_{t-d} < q_j, \tag{2.4}$$

[1]For an extensively discussion of this model and its statistical properties, see Tong (1990).

with j = 1,2,...,l and $-\infty = q_0 < q_1 < ... < q_{l-1} < q_l = +\infty$ as the thresholds, α_j denotes an intercept and p_j is the lag order of the jth regime. $d \in (1, ..., \overline{d})$ is called the delay parameter, where $\overline{d}$ is typically equal to p_{max}. e_t^j are white noise sequences, conditional upon the history of the time series I_{jt}, with zero mean $E_t[e_t^j | I_{t-1}] = 0$ and variance $E_t[e_t^2 | I_{t-1}]=\sigma_j^2$. σ_j^2s have to be mutually independent for different regimes. Threshold parameters q_j divide the sample into l piecewise linear AR(p_j) processes, conditional on a specific past value of the time series y_{t-d} and threshold value q_j. The overall process is nonlinear if at least two regimes exist.

For known parameters d, q_j and p_j, the Threshold model can easily be estimated by Ordinary Least Squares. In this case, the data is separated into its l regimes and the least squares estimate is computed for each regime individually. Nevertheless, parameters d, q_j and p_j are normally unknown and have to be estimated a-priori via a grid search. For all possible combinations of delay parameters $d \in \{1, \overline{d}\}$, threshold values $q_j \in (y_{(1)}, ..., y_{(n)})$ and lag orders $p_j \in (1, ..., p_{max}^j)$, it is straightforward to estimate a SETAR($l, 1, .., p_l$) model and compute a specific information criterion of interest. Again, a widely used criterion for this nonlinear model is Akaikes AIC. Its model selection procedure will be explained in the following.

Let $y_{(1)} \leq y_{(2)} \leq ... \leq y_{(n)}$ denote a ordered time series of interest and let d and y_{t-d} be fixed. For each fixed combination of $d \in (1, ..., \overline{d})$, $y_{t-d} \in (y_{(1)}, ..., y_{(n)})$ and a given value of $p_j \in (1, ..., p_{max}^j)$ compute the corresponding AIC for each regime separately. The selected lag orders of regimes j = 1,..,l minimize the regime-specific criterion: $\hat{p_j} = \min_{0 \leq p_i \leq p_{max}} AIC_j(p_i)$, with

$$AIC_j(p_i) = n_j log(\hat{\sigma}_j^2) + \frac{2(p_i + 1)}{n_j}. \tag{2.5}$$

n_j is the number of observations of regime j and $\hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{t=0}^{n_j} (e_t^j)^2$ is the estimated residual variance of regime j. Following Tong (1990), the AIC criterion for a Threshold model is given by the sum of regime-specific AICs[2]:

$$AIC(\hat{p_j})^{TAR} = AIC_1(\hat{p_1}) + AIC_2(\hat{p_2}) + ... + AIC_l(\hat{p_l}). \tag{2.6}$$

Next, keeping $\hat{p_j}$ for all regimes and the delay parameter d fixed, the estimated threshold value $\hat{q_j}$ is obtained by minimizing the information criterion over a possible set of threshold values:

$$\hat{q_j} = \min_{\{q_j\}} AIC(\hat{p_j})^{TAR}. \tag{2.7}$$

[2]This representation is only feasible under the assumption of e_t^j (j = 1,...,l) being mutually independent for all regimes.

Finally, keeping all $\hat{p_j}$ and $\hat{q_j}$ fixed, a search for the lowest information criterion value gives an appropriate estimate for the delay parameter $\hat{d}$:

$$\hat{d} = minAIC(\hat{d}) = \frac{AIC(\hat{q}_j)}{n - max(d, L)}. \tag{2.8}$$

$L = max(p_j)$ is the maximum lag order over all regimes. Analogous to this, all parameters can be estimated by Schwarz‘s information criterion (BIC). Nevertheless, both estimation procedures require a sufficient number of observations in each regime. Accordingly, it can be necessary to restrict the grid search to a subset of ordered observations. Andrews (1993) suggests the following interval limits for a ordered time series subset: $\pi_1 = .15$ and $\pi_2 = .85$. Using only this range for possible threshold values it is guaranteed that every regime has a minimum number of observations and a reliable estimation can be computed. If the grid search leads to an estimated threshold equal to the first value of the ordered time series, this trimming procedure ensures that even the first regime contains at least 15 percent of ordered observations. Teräsvrita (2005) argues that nonlinear models have a good chance to outperform linear models if a sufficient number of observations are available. The possible failure of nonlinear models may be due to too little observations for specifying the model and estimating its parameters.

3 Data

The following empirical application uses a huge data set of circa 40 monthly time series for each of ten European countries. All time series span from 1996:3 up to 2008:12, they are seasonally adjusted and can be classified into five groups of variables: Industrial Production Index, Consumer Price Index, Producer Price Index, Unemployment and Financial Market[3]. To obtain stationary processes, all time series are subjected to two transformations. First, all series were transformed by taking the logarithm. Second, depending on the result of an Augmented Dickey Fuller test, time series were differentiated. After taking the logarithm no time series were indicated to be stationary and hence at least one difference needed to be taken. According to Marcellino et al. (2006), absolute values that exceeded its median by more than six times its Interquartile Range, were treated as outliers. In order to avoid such defined outliers to affect the forecasting results, they were dropped. Table A.2 in the appendix lists the number of outliers and differentiations for each available time series. A graphical investigation of all time series showed a stationary fluctuation around a nonzero value with no trending behavior. Additionally, a AIC and

[3] A complete list of time series, economies and additional informations are given in the Data appendix A.

BIC search over different type of models led to no evidence of a linear trend. Therefore, the linear Autoregressive model and the nonlinear Threshold model include a constant term but no linear trend. Hereafter, $y_{i,r,t}$ is referred to a fully transformed and adjusted time series, where i=1,...,10 denotes the number of economy, r=1,...,R_i is the number of time series and t=1,...,N indicates the time index. R_i is the total number of time series of economy i.

Several studies find evidence for nonlinearity of economic variables like unemployment rates and Industrial Production indices. Therefore, all time series are tested for nonlinearity by means of two nonlinearity tests (see Keenan, 1985 and McLeod & Li, 1983). Each group of variables contains a minority of variables that are detected to be 'nonlinear', whereby the following two subsamples comprise the most detected 'nonlinear' time series: "Consume Price Index" and "Industrial Production Index". Furthermore, a test for Threshold nonlinearity is also applied (see Hansen, 1999). According to this test and its results, solely Threshold models with two regimes are used in this study.

4 Empirical application

4.1 Methods and Parametrization

As mentioned in section 2, lag orders are determined by AIC and BIC. Transformations of lagged time series values or residuals are irrelevant for both model selection procedures. Furthermore, the usage of such criterions requires a choice of a maximum lag order p_{max}. Depending on the monthly frequency of the data, a maximum lag order of 12 is applied in this study[4]. Examining the nonlinear Threshold model it was striking that larger lag orders (12 and higher) led to unreasonable high loss function values. Therefore and provided by common literature, this model is used along with a maximum lag order of six (for example, see Byers & Peel, 1995 and Clements & Smith, 1999). The determination of the 'optimal' lag order p_{max} requires a truncation of each time series. Dropping the first twelve observations for all time series, guarantees that every implementation uses the same set of information. For the purpose of an appropriate number of observations for the estimation and a sufficient quantity of forecast errors, the forecast horizon was chosen to be 23 months. Hence, the in-sample period for the first regression step spans from 1997:3 to 2007:1 and merges 119 observations. The out-of-sample period is covering the time from 2007:2 up to 2008:12.

Model based forecasts and lag order selection are computed recursively. This means, that forecasts are based on values of the time series up to the date on which the forecast is made. Only actually

[4] Additional maximum lag orders (6 and 18) have been examined but did not provide any deviating forecasting results since both information criterions usually did not select significant deviating lag orders.

available informations are used for each out-of-sample forecast. For the next forecasting step, the lag order is chosen again and parameters are reestimated. Thus, selected lag orders and estimated coefficients can vary across time. Moreover, a distinction between a rolling estimation window of fixed size ω and an expanding estimation window is made. Using a rolling estimation window, the one-step ahead forecast is added to the data set while the first observation is dropped. In this case, every forecasting step applies a fixed window size $\omega = 119 - i_0^r$ for the estimation. i_0^r is the number of outliers of time series r. Adapting an expanding estimation window, the one-step ahead forecast is added to the time series and no observation is dropped. Thus, the estimation window increases with every forecasting step. Rolling estimation windows are a useful tool for time series with structural breaks, since this estimation procedure accommodates the possible instability of AR parameters over time. Such instability leads to forecast uncertainty and it can be preferable not to use the full data set (Peseran and Timmermann, 2004). Expanding estimation windows lead to more efficient estimates. This approach is exploiting more available sample information and a steadily increasing information set can lead to a reduced estimation uncertainty (Herwartz, 2010a). This estimation method is optimal in the presence of no structural breaks in the data (Peseran and Timmermann, 2007).

As outlined in the introduction, the focus of this empirical application is on the predictive content of nonlinear transformations of lagged residuals e_t and lagged time series values y_t. Therefore, six different nonlinear transformations will be used and compared in this study: square function $(\cdot)^2$, cubic function $(\cdot)^3$, sine function $\sin(\cdot)$, cosine function $\cos(\cdot)$, tangents function $\tan(\cdot)$ and exponential function $\exp(\cdot)$. Assuming that the first lag of transformed time series contains the main predictive content[5], only the first transformed lag is added to an Autoregressive process (transformed AR(p) model):

$$y_{i,rt} = \alpha + \beta_1 y_{i,r,t-1} + \beta_2 y_{i,r,t-2} + ... + \beta_p y_{i,r,t-p} + \beta_{p+1} y^{\square}_{i,r,t-1} + e_{i,r,t}, \quad t = 1, ..., N. \tag{4.1}$$

$y^{\square}_{t-1}$ is referred to a transformed time series and is representing one of the six transformations described above. Transformations of lagged residuals and the simple ARMA(p,q) model are combined as follows (transformed ARMA(p,q) model):

$$\tilde{y}_{i,r,t} = \beta_1 \tilde{y}_{i,r,t-1} + \beta_2 \tilde{y}_{i,r,t-2} + ... + \beta_p \tilde{y}_{i,r,t-p} + \beta_{p+1} e^{\square}_{i,r,t-1} + e_{i,r,t}, \quad t = 1, ..., N. \tag{4.2}$$

Again, $e^{\square}_{i,r,t-1}$ is referred to the first lag of transformed residuals and the same assumption as before is made. Solely the first lag of transformed residuals contains important information and is able to systemat-

[5]Autoregressive models with nonlinear transformations of higher lags of time series values led to equal result conclusions and thus contain a negligible predictive content.

ically improve the forecasting performance of a simple linear model.[6] Both transformed Autoregressive models can be estimated by Ordinary Least Squares methods by simply adapting the common regression matrices.

4.2 Forecast evaluation

Once, all parameters are estimated it is easy to compute a one-step ahead forecast by means of the following two equations:

$$\hat{y}_{t+1|t} = \hat{\alpha} + \hat{\beta}_1 y_{t+1-1} + \hat{\beta}_2 y_{t+1-2} + ... + \hat{\beta}_p y_{t+1-p} + \hat{\beta}_{p+1} y^{\square}_{t+1-1} \tag{4.3}$$

and

$$\hat{y}_{t+1|t} = \hat{\alpha} + \hat{\beta}_1 y_{t+1-1} + ... + \hat{\beta}_p y_{t+1-p} + \hat{\beta}_{p+1} e^{\square}_{t+1-1}, \tag{4.4}$$

with $y_i = \tilde{y}_i + \bar{y}$ for the latter model. Compared to a multi-period ahead forecast, forecasting a nonlinear model one period ahead does not pose any problem. For example, consider a first order SETAR model with delay order one, lag orders one, threshold variable q_j and two regimes:

$$y_t = (\beta_0^1 + \beta_1^1 y_{t-1} + e_t^1) I(y_{t-1} < q_j) + (\beta_0^2 + \beta_1^2 y_{t-1} + e_t^2) I(y_{t-1} \geq q_j), \tag{4.5}$$

where $e_t^j \sim nid(0, \sigma_j^2)$, j=1,2. $I(\cdot)$ is an indicator function that is equal to one if the condition in parenthesis holds. Otherwise it is zero. The one-step ahead forecast for this SETAR model is then given by:

$$\hat{y}_{t+1|t} = E(y_{t+1}|y_t < q_j) I(y_t < q_j) + E(y_{t+1}|y_t \geq q_j) I(y_t \geq q_j), \tag{4.6}$$

where $E(y_{t+1}|y_t < q_j) = \hat{\beta}_0^1 + \hat{\beta}_1^1 y_t$ and $E(y_{t+1}|y_t \geq q_j) = \hat{\beta}_0^2 + \hat{\beta}_1^2 y_t$. One-step ahead forecasts of SETAR models with higher lag orders and further regimes are straightforward.

The following remarks have to be considered:

(1) $\hat{y}_j = y_j$ for $j \leq t$

(2) Unobservable observations y_{t+1} are replaced by its optimal forecasts $\hat{y}_{t+1}$

(3) Residuals in the first forecasting step are equal to zero: $e_t = 0$ for $t \leq 0$

[6] This assumptions was tested as well and could be confirmed for this study.

$\epsilon_{t+1} = y_{t+1} - \hat{y}_{t+1}$ is the corresponding one-step ahead forecast error. The forecast uncertainty is defined by the forecast variance:

$$\tilde{\sigma}_{i,r,t} = \sqrt{\sigma^2_{i,r,t}(1 + y_{i,r,t-1}(\mathbf{X}'\mathbf{X})^{-1}y_{i,r,t-1})}, \tag{4.7}$$

with

$$\sigma^2_{i,r,t} = \frac{1}{N-K}\sum_{t=1}^{T}\hat{e}_{i,r,t}\hat{e}_{i,r,t}. \tag{4.8}$$

N is the number of observations and K is the column size of the regressor matrix.

4.3 Measuring the forecasting performance

According to Herwartz (2010a, 2010b) and Marcellino et al. (2006), this subsection introduces five loss functions that are used to appraise the forecasting performance of two competing forecasting models. Moreover, Mean group statistics for aggregating across time series and economies are explained. Each implementation compares a specific forecasting model of interest with a benchmark model. The basic benchmark model is a linear Autoregressive model as in equation (2.1) which does not include any nonlinear transformation of lagged residuals or time series values. Based on the choice of AIC or BIC and rolling or expanding estimation window, the benchmark model uses the same specification and estimation strategy. The objective of this study are nonlinear transformations and its predictive content, different estimation methods and model selection procedures are used for robustness reasons. In order to detect a reliable statement about the average forecasting performance of transformed Autoregressive models, the benchmark model is adjusted according to selected estimation and modeling procedures. The benchmark model is labeled by * and a specific forecasting model of interest by $^\bullet$. All time series were separated into an in-sample period ($t = 13, ..., T$), and an out-of-sample period ($t = T+1, ..., N$), where T=131-i^r_o and N=154-i^r_o (see section 4.1). Each out-of-sample forecasting step provides a one-step ahead forecasting error ϵ_{t+1} and a forecasting variance $\tilde{\sigma}_{i,r,t}$ on which basis the following five loss functions are computed.

(1) Differential of relative Mean Absolute forecast Error ($DMAE$)

$$DMAE^\bullet_{i,r} = RMAE^\bullet_{i,r} - RMAE^*_{i,r} \tag{4.9}$$

with

$$RMAE^{\bullet}_{i,r} = \frac{1}{23}\sum_{t=T+1}^{N}\frac{|y_{i,r,t}-\hat{y}_{i,r,t}|}{\hat{\sigma}_{i,r,t}}. \tag{4.10}$$

$RMAE$ is the Relative Mean Absolute forecast Error and

$$\hat{\sigma}_{i,r,t} = \sqrt{\frac{1}{t-K}\sum_{j=1}^{t}\hat{e}_{i,r,j}\hat{e}_{i,r,j}} \tag{4.11}$$

is a strategy- and transformation invariant estimator of the residual variance. $\hat{e}_{i,r,t}$ are computed based on the whole set of regressors $\mathbf{X} = \{\mathbb{1}, y_{-1}, ..., y_{-p_{max}}\}$, where $\mathbb{1}$ is a constant vector of ones. K is the column size of the regressor matrix $\mathbf{X}$.

(2) Differential of frequencies for Minimum absolute forecast errors ($DMIN$)

$$DMIN^{\bullet}_{i,r} = \frac{1}{23}\sum_{t=T+1}^{N} I(|\epsilon^{*}_{i,r,t}| \leq |\epsilon^{\bullet}_{i,r,t}|) - I(|\epsilon^{\bullet}_{i,r,t}| \leq |\epsilon^{*}_{i,r,t}|), \tag{4.12}$$

$\epsilon_{i,r,t} = y_{i,r,t}$ - $\hat{y}_{i,r,t|t}$ is the forecast error and I($\cdot$) as an indicator function.

(3) Differential of frequencies for minimum ex-ante uncertainty ($DPUC$)

$$DPUC^{\bullet}_{i,r} = \frac{1}{23}\sum_{t=T+1}^{N} I(\tilde{\sigma}^{*}_{i,r,t} \leq \tilde{\sigma}^{\bullet}_{i,r,t}) - I(\tilde{\sigma}^{\bullet}_{i,r,t} \leq \tilde{\sigma}^{*}_{i,r,t}), \tag{4.13}$$

with $\tilde{\sigma}_{i,r,t}$ as the estimated ex-ante forecast uncertainty (see 4.7).

(4) Directional Accuracy loss statistic (DA)

$$DA_{i,r} = \frac{1}{23}\sum_{t=T+1}^{N} I(|\tilde{da}^{\bullet}_{i,r,t}| > |\tilde{da}^{*}_{i,r,t}|) - I(|\tilde{da}^{\bullet}_{i,r,t}| < |\tilde{da}^{*}_{i,r,t}|), \tag{4.14}$$

with

$$\tilde{da}_{i,r,t} = I(y_{i,r,t} \times \hat{y}_{i,r,t} \geq 0) - I(y_{i,r,t} \times \bar{y}_{i,r,t} \geq 0) \tag{4.15}$$

as the directional accuracy excess over the naive forecast $\bar{y}_{i,r,t} = \frac{1}{n}\sum_{t=1}^{n} y_{i,r,T-t+1}$. n is the number of observations. This forecast is averaging the in-sample observations of a time series.

(5) Relative Mean Squared Forecast Error ($RMSFE$)

$$RMSFE_{i,r} = \frac{MSFE^{\bullet}_{i,r}}{MSFE^{*}_{i,r}} \tag{4.16}$$

with

$$MSFE_{i,r} = \frac{1}{23} \sum_{t=T+1}^{N} \epsilon^2_{i,r,t}. \tag{4.17}$$

Positive values of loss functions (1),(2),(3) are in favor of the benchmark model. Negative values are related to a better forecasting performance of the model that is under consideration. The reverse condition is true for the Directional accuracy loss statistic in (4). A value greater than one for loss function $RMSFE$ provides a better forecasting performance of the benchmark model. All five loss functions are computed for each time series (382 series), for each model selection procedure and for each estimation method.

Absolute forecast errors ϵ_{t+1} and forecast uncertainties $\tilde{\sigma}_{i,r,t}$ are scale dependent measures. This may present a problem for the aggregation over time series and economies. In order to avoid this problem, all measures are converted into scale free statistics. Calculating loss function $DMAE$, Relative Mean Absolute Errors are scale adjusted by the estimated modeling-invariant in-sample standard error. Accordingly, this measure treats large and small forecast errors in the same way. Indicator functions in (2),(3) and (4) are additional helpful tools and translate the forecasting performances into scale free statistics. A disadvantage of the $DPUC$ loss function is its dependency on the model size (for example, see Herwartz, 2010a). The forecast uncertainty $\tilde{\sigma}_{i,r,t}$ is negatively related to the column size of the regressor matrix. For further discussion on this issue, see section 5.

4.4 Mean group statistics

Considering 382 available time series and 23 forecasting steps, 8786 = 382 · 23 loss function values are computed for each implementation, each considered transformation and each estimation procedure. Because of the large number of available loss functions, this study does not focus on the forecasting performance of single time series, it rather answers the question which forecasting model performs better on average and most frequently leads to the best forecasting results.

In order to compare alternative forecasting schemes, Mean Group-statistics (MG-statistics) are evaluated according to Herwartz (2010a). The forecasting performance of economy i, averaged over its R_i

time series is given by:

$$\hat{g}_i^{\bullet} = 1/R_i \sum_{r=1}^{R_i} \hat{g}_{i,r}^{\bullet}, \tag{4.18}$$

where $\hat{g}_{i,r}^{\bullet}$ represents any of the five loss functions described above. The cross sectional Mean Group statistic is then denoted by:

$$\tilde{\triangle}_G^{\bullet} = \frac{1}{10} \sum_{i=1}^{10} \hat{g}_i^{\bullet}. \tag{4.19}$$

Furthermore, the null hypothesis H_0:$\tilde{\triangle}_G^{\bullet} = 0$ is tested against the alternative hypothesis H_1:$\tilde{\triangle}_G^{\bullet} \neq 0$. For this purpose, a standard t-ratio-test using the following test statistic is applied:

$$t = \sqrt{10} \frac{\tilde{\triangle}_G^{\bullet} - 0}{\sigma_{\tilde{\triangle}_G^{\bullet}}}. \tag{4.20}$$

Testing the significance of the RMSFE loss function (5) zero is replaced by one in this statistic. $\sigma^{\bullet}_{\tilde{\triangle}_G}$ is the standard deviation of cross sectional Mean Group statistics.

5 Empirical results

5.1 Full sample results

Table 5.1 documents all MG-statistics for the full sample evaluation of the transformed AR(p) model. Test statistics are also provided in parenthesis and bolded values indicate significance at the 5% level. The left-hand side panel provides the outcome for implementations using an expanding estimation window and the right-hand side panel the rolling estimation results. As obvious from this table, the benchmark model, a simple linear model, provides more frequently minimum absolute forecast errors ($DMIN$). Accordingly, any considered transformation of lagged time series values achieves lower absolute forecast errors than a simple Autoregressive model. But, applying transformations $(\cdot)^2$, $\cos(\cdot)$ or $\exp(\cdot)$ leads to negative values that are in favor of the transformed AR(p) model. Using an expanding estimation window these values are even significant unequal to zero. The next loss function ($DMAE$), measuring the differential of relative mean absolute forecast errors is always lower than zero. Nevertheless, these values are mostly not significant in favor of transformed models. The directional accuracy over the naive forecast $\bar{y}_{i,r}$ is constantly significant higher than zero. This implies, that transformed AR(p) models lead to superior 'forecasting signs' than the benchmark model and the naive forecast. 'Forecasting signs' in this context means that forecasts of the transformed AR(P) model exhibit the same signs as the true time series values. This conclusion holds independent of the considered estimation window. Relative mean squared forecast errors (RMSFE) indicate equal forecasting results for all competing implementations. Each value of this loss function is not significant unequal to zero. On average, adding nonlinear transformations of lagged time series values to an Autoregressive process does not result in lower mean squared forecast errors.

Results for loss function $DPUC$, regarding the ex-ante forecast uncertainty are striking. Considering the expanding estimation window first, this MG statistic is positive for implementations related to transformations $(\cdot)^2$, $\cos(\cdot)$ and $\exp(\cdot)$. Accordingly, the benchmark model achieves lower ex-ante uncertainties. The remaining implementation provide not significant negative values. Forecasting results for the rolling estimation window look quite superior. Each considered transformed AR(p) model provides a highly lower ex-ante forecast uncertainty than the benchmark model. As mentioned in section 4.3 and pointed out by Herwartz (2010a), this loss function is negatively related to the number of regressors. Therefore, the ex-ante forecast uncertainty increases if the model size (K=$\hat{p}$+1) decreases and this consequently affects the results of this loss function. Whereas the AIC tends to overestimate lag order $\hat{p}$, the BIC is known for its parsimonious lag order selection[7]. In comparison to the benchmark model, both information criterions usually select a smaller lag order for transformed AR(p) models along with the rolling estimation window and for the expanding estimation window along with the BIC criterion. Therefore, the ex-ante forecast uncertainty can be reduced by applying nonlinear transformations of lagged time series values along with the rolling estimation window and the BIC model selection approach. The distinction between the AIC and BIC model selection is due to the constantly lower selected model size for the latter criterion. Hence, it is mostly leading to superior forecast uncertainty results.

Overall, transformed AR(p) models provide superior ex-ante uncertainty ($DPUC$) and directional accuracy excess (DA) loss function values. Compared to the benchmark model, absolute and squared forecast errors can not be improved by using additional transformed time series values. With respect to these loss functions, nonlinear transformations of lagged time series do not contain helpful predictive content and are not able to significantly improve the overall forecasting performance of simple AR(p) models. Nevertheless, the overall forecasting results are superior for the expanding estimation window. This approach is using more available sample information and thus generally leads to lower loss functions values that are related to forecast errors ($DMIN, DMAE, RMSFE$). The remaining two functions

[7]This property was supported by the empirical results in this study. But still, the overall lag order selection was very similar for both model selection procedures.

Table 5.1: Results for the transformed Autoregressive AR(p) model

Transformed Autoregressive model											
		expanding estimation window					rolling estimation window				
trans.	IC	DMIN	DMAE	DPUC	DA	RMSFE	DMIN	DMAE	DPUC	DA	RMSFE
$(\cdot)^2$	AIC	-2.411	0.466	7.970	0.219	1.001	-0.465	-0.020	-82.800	0.238	0.998
		(2.8)	**(-2.83)**	(1.78)	**(2.97)**	(0.45)	(-0.14)	(-0.47)	**(-73.21)**	**(3.9)**	(-0.84)
$(\cdot)^2$	BIC	-2.835	-0.144	12.182	0.291	0.999	-0.214	-0.099	-83.212	0.312	0.997
		(-0.99)	**(-2.47)**	**(2.71)**	**(2.84)**	(-0.23)	(-0.67)	(-0.21)	**(73.03)**	**(3.34)**	(-1.29)
$(\cdot)^3$	AIC	-0.658	0.746	-4.761	0.079	1.009	-0.703	0.306	-84.520	0.101	1.007
		(1.87)	(-0.38)	(-1.19)	**(3.5)**	**(2.74)**	(0.8)	(-0.4)	**(-81.62)**	**(3.46)**	**(2.27)**
$(\cdot)^3$	BIC	-0.542	0.181	-1.499	0.113	1.006	-0.868	0.255	-85.257	0.134	1.005
		(0.52)	(-0.32)	(-0.44)	**(4.0)**	**(2.36)**	(0.69)	(-0.61)	**(-80.67)**	**(3.34)**	(1.71)
$\sin(\cdot)$	AIC	-0.658	0.746	-4.761	0.079	1.009	-0.681	0.306	-84.520	0.101	1.007
		(1.87)	(-0.38)	(-1.19)	**(3.5)**	**(2.75)**	(0.8)	(-0.39)	**(-81.62)**	**(3.46)**	**(2.27)**
$\sin(\cdot)$	BIC	-0.542	0.181	-1.499	0.113	1.006	-0.868	0.255	-85.235	0.134	1.005
		(0.52)	(-0.32)	(-0.44)	**(4.0)**	**(2.36)**	(0.69)	(-0.61)	**(-80.66)**	**(3.34)**	(1.71)
$\cos(\cdot)$	AIC	-2.411	0.466	7.970	0.219	1.001	-0.465	-0.020	-82.800	0.238	0.998
		(2.8)	**(-2.83)**	(1.78)	**(2.97)**	(0.46)	(-0.14)	(-0.47)	**(-73.21)**	**(3.9)**	(-0.83)
$\cos(\cdot)$	BIC	-2.835	-0.144	12.182	0.291	0.999	-0.214	-0.099	-83.212	0.312	0.997
		(-0.99)	**(-2.47)**	**(2.71)**	**(2.84)**	(-0.22)	(-0.67)	(-0.21)	**(-73.03)**	**(3.34)**	(-1.28)
$\tan(\cdot)$	AIC	-0.503	0.744	-4.699	0.079	1.009	-0.659	0.305	-84.520	0.101	1.007
		(1.87)	(-0.29)	(-1.19)	**(3.5)**	**(2.72)**	(0.8)	(-0.38)	**(-81.62)**	**(3.46)**	(2.25)
$\tan(\cdot)$	BIC	-0.499	0.179	-1.436	0.113	1.006	-0.868	0.254	-85.257	0.134	1.005
		(0.52)	(-0.29)	(-0.42)	**(4.0)**	**(2.35)**	(0.69)	(-0.6)	**(-80.67)**	**(3.34)**	(1.69)
$\exp(\cdot)$	AIC	-2.320	0.464	7.618	0.219	1.001	-0.577	-0.021	-82.799	0.238	0.998
		(2.78)	**(-2.63)**	(1.69)	**(2.97)**	(0.58)	(-0.14)	(-0.59)	**(-73.94)**	**(3.9)**	(-0.75)
$\exp(\cdot)$	BIC	-2.874	-0.145	11.764	0.291	1.000	-0.480	-0.099	-83.256	0.312	0.997
		(-0.99)	**(-2.5)**	**(2.78)**	**(2.84)**	(-0.15)	(-0.66)	(-0.46)	**(-73.47)**	**(3.34)**	(-1.25)

Note: MG- and test-statistics $\tilde{\triangle}^{\bullet}_{G} = 0$ for loss functions $DMIN$, $DMAE$, $DPUC$, DA and $RMSFE$. A maximum lag order of 12 is used in this application. MG statistics are multiplied by 100 (except for loss function $RMSFE$). Small numbers in parenthesis denote t-ratios for testing the null hypothesis $H_0 : \tilde{\triangle}^{\bullet}_{G} = 0$(or 1). Bolded values indicate significance at the 5% level. Considered nonlinear transformations (trans.) and information criterions (IC) are given as well.

DA and $DPUC$ provide superior forecasting results for the rolling estimation window. Furthermore, a significant distinction between the various considered transformation is not recognizable for this model. Using the expanding estimation window it is striking that transformations $(\cdot)^2$, $\cos(\cdot)$ and $\exp(\cdot)$ are leading to the best results for loss functions $DMIN$, $DMAE$ and DA. This is especially true for BIC model selection procedure. Considering a rolling estimation window these distinction are not visible anymore. Solely loss function DA is leading to superior forecasting results for these transformations.

Table 5.2: Results for the transformed Autoregressive Moving Average model ARMA(p,q)

Transformed Autoregressive Moving Average model											
		expanding estimation window					rolling estimation window				
trans.	IC	DMIN	DMAE	DPUC	DA	RMSFE	DMIN	DMAE	DPUC	DA	RMSFE
	AIC	-14.385	-0.818	-66.272	2.272	1.187	-12.118	1.676	-89.352	2.106	1.126
		(-7.24)	(-1.2)	**(-17.17)**	**(15.43)**	**(6.02)**	**(-7.86)**	(0.85)	**(-68.32)**	**(11.94)**	(-0.1)
	BIC	-16.877	-1.369	-67.463	2.223	0.982	-14.082	-1.438	-92.398	2.002	0.988
		(-8.34)	**(-2.37)**	**(-16.57)**	**(12.99)**	(-1.45)	**(-8.82)**	**(-3.22)**	**(-68.7)**	**(11.86)**	(-1.72)
$(\cdot)^2$	AIC	-16.003	-1.392	-58.243	2.269	0.980	-15.044	-1.421	-86.205	2.080	0.984
		(-10.4)	**(-3.14)**	**(-11.63)**	**(13.81)**	(-2.03)	**(-12.38)**	**(-3.36)**	**(-44.74)**	**(11.93)**	(-1.67)
$(\cdot)^2$	BIC	-16.894	-1.459	-51.645	2.277	0.980	-15.615	-1.496	-81.055	2.002	0.985
		(-8.67)	**(-3.4)**	**(-9.84)**	**(13.27)**	(-2.01)	**(-12.64)**	**(-3.78)**	**(-28.14)**	**(12.12)**	(-1.12)
$(\cdot)^3$	AIC	-15.968	-1.357	-60.114	2.269	0.983	-14.659	-1.326	-87.746	2.071	0.988
		(-7.93)	**(-2.93)**	**(-14.01)**	**(13.81)**	(-1.13)	**(-9.28)**	**(-3.15)**	**(-54.06)**	**(12.34)**	(-1.22)
$(\cdot)^3$	BIC	-17.137	-1.438	-61.577	2.255	0.982	-15.344	-1.447	-86.013	2.002	0.987
		(-8.12)	**(-3.2)**	**(-11.49)**	**(12.62)**	(-1.15)	**(-9.7)**	**(-3.45)**	**(-31.25)**	**(12.12)**	(-0.87)
$\sin(\cdot)$	AIC	-14.550	-0.898	-65.713	2.272	1.081	-12.259	3.414	-88.939	2.106	1.092
		(-7.44)	(-1.42)	**(-15.75)**	**(15.43)**	**(10.21)**	**(-7.94)**	(0.9)	**(-64.34)**	**(11.94)**	(-0.18)
$\sin(\cdot)$	BIC	-16.834	-1.311	-66.731	2.223	1.002	-14.102	-1.438	-92.376	2.002	0.988
		(-8.35)	**(-2.35)**	**(-15.45)**	**(12.99)**	**(9.52)**	**(-8.73)**	**(-3.21)**	**(-68.81)**	**(11.86)**	(-1.7)
$\cos(\cdot)$	AIC	-12.589	0.322	61.158	1.904	1.002	-11.457	0.400	-78.453	1.646	1.001
		(-6.75)	(1.59)	**(37.7)**	**(11.49)**	**(7.08)**	**(-6.11)**	(1.63)	**(-36.73)**	**(8.32)**	**(7.22)**
$\cos(\cdot)$	BIC	-14.434	0.375	77.192	1.879	1.002	-13.673	0.352	-75.876	1.601	1.002
		(-7.02)	(2.24)	**(23.09)**	**(10.48)**	**(8.97)**	**(-6.98)**	(1.98)	**(-21.43)**	**(8.09)**	**(9.16)**
$\tan(\cdot)$	AIC	-14.376	-0.949	-66.988	2.270	1.067	-12.372	0.521	-89.525	2.094	1.011
		(-7.44)	(-1.49)	**(-17.55)**	**(14.65)**	**(10.29)**	**(-8.1)**	(0.37)	**(-65.78)**	**(11.69)**	(-0.73)
$\tan(\cdot)$	BIC	-16.921	-1.361	-67.681	2.234	0.981	-14.106	-1.440	-92.398	2.002	0.988
		(-8.33)	**(-2.35)**	**(-16.49)**	**(12.9)**	**(-2.76)**	**(-8.97)**	**(-3.23)**	**(-68.7)**	**(11.86)**	(-1.85)
$\exp(\cdot)$	AIC	-12.987	0.320	41.841	1.880	1.002	-11.059	0.391	-78.499	1.635	1.000
		(-6.75)	(1.6)	**(14.41)**	**(10.7)**	**(7.01)**	**(-5.5)**	(1.61)	**(-36.8)**	**(8.54)**	**(7.15)**
$\exp(\cdot)$	BIC	-13.845	0.372	45.043	1.867	1.002	-13.136	0.347	-75.715	1.590	1.001
		(-7.2)	(2.24)	**(10.02)**	**(10.25)**	**(9.07)**	**(-7.03)**	(1.97)	**(-21.16)**	**(8.29)**	**(9.11)**

Note: MG- and test-statistics $\tilde{\triangle}^{\bullet}_{G} = 0$ for loss functions $DMIN$, $DMAE$, $DPUC$, DA and $RMSFE$. A maximum lag order of 12 is used in this application. MG statistics are multiplied by 100 (except for loss function $RMSFE$). Small numbers in parenthesis denote t-ratios for testing the null hypothesis $H_0 : \tilde{\triangle}^{\bullet}_{G} = 0$(or 1). Bolded values indicate significance at the 5% level. Considered nonlinear transformations (trans.) and information criterions (IC) are given as well.

Table 5.2 lists all loss functions values for the comparison of the benchmark model and transformed ARMA(p,q) models. The first two rows compare a simple ARMA(p,q) model with no transformed lagged residuals to a linear Autoregressive model. It is obvious, that a simple ARMA(p,q) model already represents an improvement compared to a simple Autoregressive model. Almost all loss functions are significant in favor of the ARMA(p,q) model. Applying the rolling estimation window, the ex-ante uncertainty ($DPUC$) can be greatly reduced by not applying a simple linear process but an ARMA(p,q) process. Forecast errors ($DMIN$, $DMAE$ and $RMSFE$) are as well usually lower if such a model is used. The directional accuracy over the naive forecast (DA) clearly provides significant inferior forecasting results for the benchmark model. Furthermore, nonlinear transformations of lagged residuals do carry a even greater predictive content. Any considered transformed ARMA(p,q) model mostly outperforms the linear benchmark model. Particular for the rolling estimation window, all MG-statistics are almost anytime significant in favor of transformed models. Again, the ex-ante uncertainty loss function depends on the model size and leads to superior forecasting results for the BIC model selection approach. A transformed model is leading to a lower ex-ante uncertainty if the rolling estimation window is applied. Using the expanding estimation window leads to partially positive values that are in favor of the benchmark model. Whereas the majority of loss function values for the rolling estimation procedure are significant in favor of transformed ARMA(p,q) models, forecasting results for the expanding window do not look that clear. Both, mean absolute forecast errors and mean square forecast errors are not significantly negative. A clear distinction between the forecasting results of the benchmark model and the transformed ARMA(p,q) model can not be found for these cases.

Nevertheless, the following three transformations constantly achieve the best forecasting results: $(\cdot)^2$, $(\cdot)^3$, $\sin(\cdot)$ and $\tan(\cdot)$. These transformations obtain the lowest loss function values connected to absolute and mean squared forecast errors ($DMIN$ and $DMAE$) and mainly along with the BIC model selection procedure. The ex-ante uncertainty loss function provides the best forecasting results for transformations $\sin(\cdot)$ and $\tan(\cdot)$. The remaining two loss functions (DA and $RMSFE$) provide no outstanding results for certain transformations. But still, nonlinear transformations of lagged residuals contain important predictive content and obviously improve the forecasting performance of simple Autoregressive models.

5.2 Subsample results

Considering all available time series, this study figured out that nonlinear transformations of lagged time series and lagged residuals contain a significant predictive content and help to improve the forecasting performance of simple linear Autoregressive models. Especially nonlinear transformations of lagged residuals reduce both, forecasting uncertainty and forecast errors. Certain transformations of lagged residuals led to even better results with respect to forecast errors than others. This section now examines whether these results can be carried over to different subsamples of time series. The objective is to find reliable statement about the forecasting performance of transformed Autoregressive models with respect to different economic variables and may figure out which transformed model works best for which type of time series.

The overall forecasting performance of each considered group of variables[8] is very similar to the forecasting performance of the full sample. Especially subsamples ”Industrial Production Index” (88 time series), ”Consumer Price Index” (126 time series) and ”Producer Price Index” (57 time series) which includes the most detected nonlinear time series provide the same magnitude of results. The main difference lies in the loss functions’ significance. Especially the transformed AR(p) model leads

[8]The minority of time series of each group have been tested to be nonlinear according to both considered nonlinearity tests (see section 3).

to barely deviating forecasting results for all type of considered variables. Several loss functions provide constantly higher or lower loss function, but remain the same result conclusions as the full sample evaluation.

Applying both transformed Autoregressive models for time series that have been detected to be 'nonlinear' according to nonlinearity tests leads to quite different forecasting results. Whereas Keenan's nonlinearity test detects 133 time series to be nonlinear, McLeod and Li's tests solely finds 82 nonlinear time series. Transformed AR(p) models generally provide inferior forecasting results for nonlinear time series according to Keenan‘s test. But, applying a transformed ARMA(p,q) model to these time series leads to significant better forecasting results. This is especially true for the expanding estimation window with loss function values almost as twice as big as before. Nevertheless, an outstanding nonlinear transformation is no longer recognizable. Using a transformed AR(p) model for nonlinear time series according to McLeod and Li's test provides loss function values that are strongly in favor of the benchmark model and is therefore not recommended. Transformed ARMA(p,q) models provide similar results as the full sample evaluation and no great distinction is recognizable. These general results can be carried over to subsamples of time series detected to be nonlinear. Any transformation of lagged residuals provide especially good forecasting results if time series are tested to be nonlinear according to Keenan's test. Nevertheless, the overall performance of both transformed models does not depend on considered group of variables.

5.3 Results for the (Self-Exciting) Threshold Autoregressive model

This last subsection compares the forecasting results of the nonlinear (Self-Exciting) Threshold Autoregressive model (SETAR) and the simple linear Autoregressive model. Its related forecasting results are stored in Table 5.3. Considering the full sample evaluation first, the overall forecasting performance is clearly in favor of the nonlinear SETAR model. Loss function $DMAE$ provides negative values that are significant in favor of the nonlinear model (expanding estimation window). Applying a rolling estimation window this is only true for the BIC model selection approach. The ex-ante uncertainty can be significantly reduced by applying a Threshold model using a rolling estimation window. Applying an expanding estimation window does not lead to a significant higher forecast uncertainty for the benchmark model. With respect to the directional accuracy excess over the naive forecast (DA) nonlinear Threshold models are again preferable. Carrying a value around two, this loss function clearly indicates better forecasting results for the nonlinear model than for the benchmark model. Regarding Relative Mean Squared Forecast errors ($RMSFE$), the linear benchmark model provides significant lower forecast errors for

Table 5.3: Results for (self-Exciting) Threshold Autoregressive model

Threshold Autoregressive Model										
	expanding estimation window					rolling estimation window				
IC	DMIN	DMAE	DPUC	DA	RMSFE	DMIN	DMAE	DPUC	DA	RMSFE
All time series										
AIC	16.154	-82.057	12.118	2.075	1.089	16.597	16.098	-42.279	2.212	1.0817
	(8.78)	**(-36.99)**	(1.82)	**(16.66)**	**(3.27)**	**(8.96)**	**(6.44)**	**(-8.46)**	**(16.27)**	**(2.89)**
BIC	15.524	-82.648	1.226	2.116	1.083	15.684	-83.103	-49.725	2.055	1.043
	(9.96)	**(-36.46)**	(0.28)	**(28.25)**	(1.31)	**(11.24)**	**(-37.85)**	**(-12.87)**	**(16.26)**	(2.17)
Industrial Production Index										
AIC	-140.095	-0.593	24.802	1.433	1.133	-140.920	0.593	-19.565	1.433	1.137
	(-0.22)	**(-25.98)**	**(3.00)**	**(6.58)**	**(3.07)**	(0.22)	**(-26.18)**	**(2.88)**	**(6.58)**	(1.43)
BIC	-141.318	-4.644	10.97	2.519	1.082	-142.221	-2.569	-36.561	2.519	1.086
	(-1.93)	**(-25.52)**	(1.63)	**(11.01)**	**(2.79)**	(-1.10)	**(-25.69)**	**(-7.13)**	**(11.01)**	**(2.91)**
Consumer Price Index										
AIC	-100.215	3.244	-0.345	2.519	1.159	-100.740	1.311	-44.859	2.519	1.161
	(1.16)	**(-17.07)**	(-0.05)	**(13.17)**	**(4.56)**	(0.48)	**(-17.17)**	**(-9.26)**	**(13.17)**	**(4.76)**
BIC	-101.222	3.658	-12.008	1.967	1.186	-101.668	2.346	-62.526	1.967	1.190
	(1.25)	**(-15.89)**	(-1.96)	**(10.20)**	**(3.58)**	(0.84)	**(-16.02)**	**(-16.79)**	**(10.20)**	**(3.73)**
Producer Price Index										
AIC	-18.665	37.757	18.993	2.212	2.035	-17.467	36.994	-44.012	2.212	2.024
	(11.37)	**(-2.81)**	(1.86)	**(7.68)**	**(4.50)**	**(10.93)**	(-2.67)	**(-5.47)**	**(7.68)**	**(4.45)**
BIC	-20.397	38.062	38.673	1.526	1.943	-19.146	37.452	-38.825	1.526	1.932
	(12.50)	**(-3.29)**	**(5.41)**	**(5.55)**	**(4.85)**	**(12.51)**	**(-3.16)**	**(-4.77)**	**(5.55)**	**(4.79)**
Unemployment										
AIC	31.544	-59.786	11.192	2.035	1.454	30.435	23.040	-45.236	2.405	1.451
	(8.47)	**(-6.31)**	(1.08)	**(6.43)**	(2.55)	**(8.51)**	**(5.9)**	**(-7.02)**	**(7.63)**	(2.50)
BIC	26.179	-62.397	10.453	2.220	1.384	26.919	-60.605	-50.416	2.590	1.378
	(6.15)	**(-6.63)**	(0.91)	**(7.00)**	**(2.91)**	**(7.04)**	**(-6.48)**	**(-7.48)**	**(8.32)**	**(2.83)**
Financial Market										
AIC	-37.229	38.440	13.596	2.001	1.467	-39.612	38.026	-43.133	2.001	1.461
	(11.00)	**(-4.77)**	(1.46)	**(7.33)**	**(4.10)**	**(10.72)**	**(-5.22)**	**(-7.25)**	**(7.33)**	**(4.0)**
BIC	-33.252	42.443	3.244	1.725	1.449	-35.689	41.477	-51.691	1.725	1.444
	(12.28)	**(-4.19)**	(0.40)	**(6.44)**	**(4.35)**	**(11.99)**	**(-4.74)**	**(-9.26)**	**(6.44)**	**(4.19)**

Note: MG- and test-statistics $\tilde{\triangle}^{\bullet}_{G} = 0$ for loss functions $DMIN$, $DMAE$, $DPUC$, DA and $RMSFE$. A maximum lag order of 6 is used in this application. MG statistics are multiplied by 100 (except for loss function $RMSFE$). Small numbers in parenthesis denote t-ratios for testing the null hypothesis $H_0 : \tilde{\triangle}^{\bullet}_{G} = 0$(or 1). Bolded values indicate significance at the 5% level. Considered information criterions (IC) are given as well.

the AIC criterion. Nevertheless, the benchmark model more frequently provides lower absolute forecast errors ($DMIN$).

Considering subsamples of time series next, these forecasting results look almost the same. Especially subsample "Unemployment" achieves the same magnitude of loss functions. Time series related to groups "Industrial Production Index" and "Consumer Price Index" provide forecasting results that are even stronger in favor of the nonlinear SETAR model. It is striking that loss function $DMIN$ provides highly negative but still not significant values. Loss function $DMAE$ exhibits positive values that are significant in favor of the linear benchmark model. The remaining loss functions provide the same results as for the full sample evaluation. Subsamples "Producer Price Index" and "Financial Market" are leading to superior values of loss function $DMIN$.

6 Conclusion

This study examined whether nonlinear transformation of lagged residuals or time series values carry important predictive content that improves the average forecasting performance of simple Autoregressive models. Furthermore, it investigated the forecasting performance of a simple nonlinear model, the (Self exciting) Threshold Autoregressive model. A large scale comparison over 382 time series from ten European economies was applied. The forecasting performance was appraised by means of several loss functions, Mean Group statistics and simple t-test statistics. Each implementation compared a specific forecasting model with a benchmark model, the simple Autoregressive model. Furthermore, all models have been tested for its robustness over different types of economic variables. Three notably findings can be detected from this empirical application.

The first major finding is that nonlinear transformations of lagged residuals (transformed ARMA(p,q) models) provide a mostly significant better forecasting performance than the benchmark model. Estimating transformed ARMA(p,q) models by a rolling window procedure leads to further enhancements compared to forecast models estimated by an expanding window. The best forecasting results constantly appeared in conjunction with the following three transformations of lagged residuals: $(\cdot)^2$, $(\cdot)^3$ and $\cos(\cdot)$.

A second main finding is that Autoregressive models with additional transformations of lagged time series values (transformed AR(p) model) do not generally lead to superior forecasting results. Absolute and mean squared forecast errors can not be reduced by transformations of lagged time series. Nevertheless, the ex-ante forecast uncertainty (measured by minimum estimates of the forecast errors standard deviation) is lower for transformed AR(p) models, as well as the loss function measuring directional accuracy excess over a naive forecast ($\bar{y}_{i,r}$). A significant distinction between various nonlinear transformations can not be found. These results are in particular true for the expanding estimation window. In addition, subsample implementations have been evaluated to investigate whether these findings are robust for different types of economic variables (Industrial Production Index, Consumer Price Index, Producer Price Index, Unemployment and Financial Market). The previous described results can be carried over to each considered group of variables. There is no clear result that transformed Autoregressive models most frequently performs the best for certain type of time series. Furthermore, it can not useful to apply transformed models especially to time series that have been detected to be nonlinear according to nonlinearity tests. This procedure leads to significant inferior forecasting results for certain type of tests.

The last finding is that a nonlinear Threshold model is generally able to capture certain behavior in the data and therefore provides better forecasting results than a simple Autoregressive model, especially for the rolling estimation window. This finding is widely independent of considered groups of time series. Nevertheless, such a nonlinear Threshold model is unsuitable for the usage with higher lag orders and should be applied along with relative low lag orders. Furthermore, applying an AIC or BIC procedure for the lag order selection of all considered models does not lead to significant deviating forecast results since both criterion usually choose similar model sizes.

A Data appendix

A total of 382 time series have been examined in this study. All time series were categorized into five groups of variables: Industrial Production Index (88 series), Consumer Price Index (126 series), Producer Price Index (57 series), Unemployment (48 series) and Financial Market (63 series). They originate from different sources like Eurostat, IMF International Financial Statistics or Main Economic Indicators by the OECD and are online available via Datastream.

Table A.1: Detailed data information

no.	series	description	information
		Industrial Production Index	
1	ips11	Industrial production index: total index	2000=100, pc
2	ips12	Industrial production index: consumer goods	2005=100, vc
3	ips13	Industrial production index: consumer durable goods	2005=100, vc
4	ips25	Industrial production index: manufacturing - electrical equipment	2005=100, vc
5	ips43	Industrial production index: manufacturing, total	2005=100, vc
6	ipi	Industrial production index: intermediate goods	2005=100, vc
7	ipmi	Industrial production index: mining and quarrying	2005=100, vc
8	mdq	Industrial production index: new orders, capital goods	2005=100, vc
9	moq	Industrial production index: new orders, manufacturing	2005=100, vc
		Financial Market	
10	fy	Long term government Bond yield in	in %
11	fm1	Money supply M1: Notes and coins in circulation, traveler's checks of non-bank issuers, demand deposits, other checkable deposits	million €/£/krona
12	fm2	Money supply M2: M1 + savings deposits, time deposits	million €/£/krona
13	fm3	Money supply M3: M2 + large time deposits	million €/£/krona
14	ex	Nominal effective exchange rate: cpi based, real	2005=100, vc
15	fer	Foreign exchange rate reserves	million US $
16	spi	Share price index	2000=100
		Unemployment	
17	lu	Total unemployment rate	in %
18	lu25	Unemployment rate: persons under 25	in %
19	lo25	Unemployment rate: persons over 25	in %
20	luw	Unemployment rate: women	in %
21	lum	Unemployment rate: men	in %
		Producer price index (PPI)	
22	pw	PPI: consumer goods	2005=100, pc
23	pwx	PPI: consumer goods excluding food, beverages and tobacco	2005=100, pc
24	pwd	PPI: durable consumer goods	2005=100, pc
25	pwn	PPI: non durable consumer goods	2005=100, pc
26	pwi	PPI: intermediate goods	2005=100, pc
27	pwm	PPI: manufacturing	2005=100, pc
		Consumer price index (CPI)	
28	pu	CPI: all items (harmonized)	2005=100, pc
29	pu882	CPI: industrial goods	2005=100, pc
30	pu81	CPI: food and non alcoholic beverages	2005=100, pc
31	pu83	CPI: clothing and footwear	2005=100, pc
32	pu84	CPI: transport	2005=100, pc
33	puh	CPI: housing, water electricity, gas and other fuels	2005=100, pc
34	pus	CPI: miscellaneous goods and services	2005=100, pc
35	puc	CPI: communications	2005=100, pc
36	pux	CPI: all items less seasonal food	2005=100, pc
37	puxh	CPI: all items less housing, water, electricity, gas and other fuels	2005=100, pc
38	pum	CPI: all items less education, health and social protection	2005=100, pc
39	pu	CPI: all itmes less energy	2005=100, pc
40	puxs	CPI: all items less services	2005=100, pc

Note: Labels and description of all time series. All time series are seasonally adjusted and transformed into stationary time series (via logarithm and differentiation up to two times). ps denotes a price index and vc indicates a volume index. The base period is indicated by the year equal to 100.

Table A.2: Panel data description

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
	ips11	ips12	ips13	ips25	ips43	ipi	ipmi	mdq	moq	fy	fm1	fm2	fm3	ex	mssing
Denmark	x	x	x(1)	x	x	x	x	x	x	x	x	x	x	x	-
Finland	x	x	x	x	x(2)	x	x	x	x	x(2)	x(2)	x(3)	x(1)	x	-
France	x	x	x	x	x	x	x	x	x	x	x	x	x	x*	-
Germany	x	x	x	x	x	x	x	x	x	x	x(1)	x	x	x	-
Italy	-	x	x	-	x	x	x	x	x	x	x(3)	x	x	x	1,4
Netherlands	x	x	x(1)	x	x	x	x	x	x	x	x(2)	x	x	x	-
Portugal	x	x	x	x(1)	x	x	x	x	x	x	-	-	-	x	11,12,13
Spain	x	x	x	x	x	x	x	x	x	x	-	-	-	x*	11,12,13
Sweden	x	x	x	x	x(2)	x	x	x	x	x	x	-	x	x	12
United Kingdom	x	x	x	x(1)	x	x	x	x	x	x	x(1)	x	x	x	-
	15	16	17	18	19	20	21	22	23	24	25	26	27		
	fer	spi	lu	lu5	lo5	luw	lum	pw	pwx	pwd	pwn	pwi	pwm		
Denmark	x	x	x(1)	x	x	x(1)	x(2)	x(1)	x	x	x	x	x		-
Finland	x	x	x	x	x	x(2)	x	x(1)	x(1)	x	x(1)	x	x(1)		-
France	x	x	x	x	x	-	x	x	x	x	x	x	x		20
Germany	x	x	x	x	x	x(3)	x	x(1)	x	x	x	x	x		-
Italy	x*	x	x	x	x	-	x	x	x	x	x	x	x		20
Netherlands	x	x	x	x(5)	x	x	x(1)	x	x	-	x	x(2)	x		23
Portugal	x	x	x(7)	x(1)	x	x	x	x	x	x	-	-	x		24,25
Spain	x*	x	x	x	x	x(8)	x	x*	x	x	x	x	x		-
Sweden	x	x	x	x	x(1)	x	x	x(3)	x	x	x	x	x		-
United Kingdom	x	x	x(7)	x	x(3)	x(1)	x	x	x	x	x(6)	x	x		-
	28	29	30	31	32	33	34	35	36	37	38	39	40		
	pu	pu882	pu81	pu83	pu84	puh	pus	puc	pux	puxh	pum	pue	puxs		
Denmark	x(6)	x	x	x	x	x	x	x	x	x	x	x	x		-
Finland	x(1)	x	x	x	x	x	x	x	x	x	x	x	x(3)		-
France	x(3)	x	x	x	x	x	x	x	x	x	x	x	x		-
Germany	x(4)	x	x	x	x	x	x	x	x	x(1)	x(1)	x(1)	x(1)		-
Italy	x(1)	x	x	x(2)	x(3)	-	x(1)	x(3)	x	x	-	x	x		33,38
Netherlands	x	x	x	x	x	x	x	x	x	-	x	x	x		28,37
Portugal	-	x	x	x	x	x	x	x	x	x	x	x	x		-
Spain	x	x	x*	x	x*	x	x	x	x	x	x	x	x		-
Sweden	x	x	x	x	x	x	x	x	x	x(1)	x	x	x		-
United Kingdom	x(1)	x	x	x	x	x	x	x	x	x(1)	x	x	x		-

Note: Panel data for the period 1996:03 up to 2008:12. Labels and numbers of time series are explained in Table A.1. Entries 'x' denote a respective series that is contained in the data set. Numbers in parenthesis indicate the number of outliers i_o. Time series labeled by * are twice differentiated. All other time series are differentiated once.

References

Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, *61*(4), 821–856.

Byers, J., Peel, D. (1995). Forecasting industrial production using non-linear methods. *Journal of Forecasting*, *14*(4), 325–336.

Clements, M., Smith, J. (1999). A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics*, *14*, 123–141.

Diebold, F., Nason, J. (1990). Nonparametric exchange rate prediction? *Journal of International Econometrics*, *2,8*(3-4), 315–332.

Granger, C., Newbold, P. (1976). Forecasting Transformed Series. *Journal of the Royal Statistical Society B*, *38*(2), 189–203.

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

Herwartz, H. (2010a). Forecast accuracy and uncertainty in applied econometrics: a recommendation of specific-to-general predictor selection. *Empirical Economics*, (pp. 1–24).

Herwartz, H. (2010b). Predictor search in autoregressive distributed lag models - an empirical assessment.

Jones, R. (1975). Fitting Autoregressions. *Journal of the American Statistical Association*, *70*(351), 590–592.

Keenan, D. M. (1985). A Tukey Nonadditivity-Type Test for Time Series Nonlinearity. *Biometrika*, *72*(1), 39–44.

Lütkepohl, H., Xu, F. (2009). The role of the log transformation in forecasting economic variables. CESifo Working Paper Series 2591, CESifo Group Munich.

Marcellino, M., Stock, J. H., Watson, M. W. (2006). A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics*, *135*, 499–526.

McLeod, A., Li, W. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, *5*(4), 269–273.

Ohtani, K. (2003). Selection of Independent Variables in a Linear Regression Model with Autocorrelated Errors: Some Small Sample Properties. *Kobe University Economic Review*, *49*, 45–50.

Peseran, M., Timmermann, A. (2004). How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting*, *20*(3), 411–425.

Peseran, M., Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, *137*(1), 134–161.

Schlittgen, R., Streitberg, B. (2001). *Zeitreihenanalyse*. Munich, Vienna: R. Oldenbourg Verlag.

Teräsvrita, T. (2005). Forecasting economic variables with nonlinear models. *SSE/EFI Working Paper Series in Economics and Finance*, *598*.

Teräsvrita, T., van Dijk, D., Medeiros, M. C. (2003). Smooth transition autoregressions, neural networks, and linear models in forecasting macroeconomic time series: A re-eximination. *International Journal of forecasting*, *21*, 755–774.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford: Clarendon Press.

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

HWWI Research Papers

since 2011

112 Regional labor demand and national labor market institutions in the EU15
Helmut Herwartz, Annekatrin Niebuhr, Hamburg, October 2011

111 Unemployment Duration in Germany – A comprehensive study with dynamic hazard models and P-Splines
Torben Kuhlenkasper, Max Friedrich Steinhardt, Hamburg, September 2011

110 Age, Life-satisfaction, and Relative Income
Felix FitzRoy, Michael Nolan, Max Friedrich Steinhardt, Hamburg, July 2011

109 The conjoint quest for a liberal positive program: "Old Chicago", Freiburg and Hayek
Ekkehard Köhler, Stefan Kolev, Hamburg, July 2011

108 Agglomeration, Congestion, and Regional Unemployment Disparities
Ulrich Zierahn, Hamburg, July 2011

107 Efficient Redistribution: Comparing Basic Income with Unemployment Benefit
Felix FitzRoy, Jim Jin, Hamburg, March 2011

106 The Resource Curse Revisited: Governance and Natural Resources
Matthias Busse, Steffen Gröning, Hamburg, March 2011

105 Regional Unemployment and New Economic Geography
Ulrich Zierahn, Hamburg, March 2011

104 The Taxation-Growth-Nexus Revisited
K. P. Arin, M. Berlemann, F. Koray, T. Kuhlenkasper, Hamburg, January 2011

The **Hamburg Institute of International Economics (HWWI)** is an independent economic research institute, based on a non-profit public-private partnership, which was founded in 2005. The University of Hamburg and the Hamburg Chamber of Commerce are shareholders in the Institute .

The HWWI's main goals are to:
• Promote economic sciences in research and teaching;
• Conduct high-quality economic research;
• Transfer and disseminate economic knowledge to policy makers, stakeholders and the general public.

The HWWI carries out interdisciplinary research activities in the context of the following research areas:
• Economic Trends and Global Markets,
• Regional Economics and Urban Development,
• Sectoral Change: Maritime Industries and Aerospace,
• Institutions and Institutional Change,
• Energy and Raw Material Markets,
• Environment and Climate,
• Demography, Migration and Integration,
• Labour and Family Economics,
• Health and Sports Economics,
• Family Owned Business, and
• Real Estate and Asset Markets.

Hamburg Institute of International Economics (HWWI)

Heimhuder Str. 71 | 20148 Hamburg | Germany
Phone: +49 (0)40 34 05 76 - 0 | Fax: +49 (0)40 34 05 76 - 776
info@hwwi.org | www.hwwi.org