Data Mining for the Detection of Turning Points in Financial Time Series

Thorsten Poddig and Claus Huber

University of Bremen, Chair of Finance
FB 7, Hochschulring 40
D-28359 Bremen, Germany
poddig@uni-bremen.de
huberc@uni-bremen.de

Abstract. One of the most challenging problems in econometrics is the prediction of turning points in financial time series. We compare ARMA- and Vector-Autoregressive (VAR-) models by examining their abilities to predict turning points in monthly time series. An approach proposed by Wecker[1] and enhanced by Kling[2] forms the basis to explicitly incorporate uncertainty in the forecasts by producing probabilistic statements for turning points. To allow for possible structural change within the time period under investigation, we conduct Data Mining by using rolling regressions over a fix-sized window. For each datapoint a multitude of models is estimated. The models are evaluated by an economic performance criterion, the Sharpe-Ratio, and a testing procedure for its statistical significance developed by Jobson/Korkie[3]. We find that ARMA-models seem to be valuable forecasting tools for predicting turning points, whereas the performance of the VAR-models is disappointing.

1 Introduction

Facing the task of forecasting with a quantitative model, an economist usually estimates a single model to produce point forecasts. Thereby the uncertainty inherent in any kind of forecast is neglected: "The generation of a forecast is of no great practical value if some measure of the uncertainty of that forecast cannot also be provided."[4]. In this paper, our intention is to explicitly incorporate this uncertainty into probabilistic statements for turning points in monthly financial time series. We implement a Monte-Carlo-based regression introduced by Wecker[1] and enhanced by Kling[2], which is described in section 2. To decide which models (ARMA or VAR) perform better we take the view of a participant in the financial markets. Here one is not interested to optimize statistical criteria, like Mean Squared Error etc., but in an acceptable profit for the taken risk. The Sharpe-Ratio is a performance criterion which allows to relate the profits to the risk of an investment. In section 3 we briefly review its basics and a test for statistical significance of the Sharpe Ratio of two investments. Section 4 describes our Data Mining approach to out-of-sample model selection

D.J. Hand, J.N. Kok, M.R. Berthold (Eds.): IDA'99, LNCS 1642, pp. 427–436, 1999.
© Springer-Verlag Berlin Heidelberg 1999

with the rolling regressions and the research design to compare the ARMA- and VAR-models. Section 5 presents empirical results and concludes.

2 Probabilistic Statements for Turning Points in Time Series

As a first step to obtain a probabilistic statement about a near-by turning point one has to define a rule when a turning point in the time series is detected. The turning point indicator

$$z_t^P = \begin{cases} 1, \text{ if a peak occurs at time } t \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

is defined as a local extreme value of a certain amount of preceding and succeeding datapoints:

$$z_t^P = \begin{cases} 1, \text{ if } x_t > x_{t+i}, i = -\tau, -\tau+1, \ldots, -1, 1, \ldots, \tau-1, \tau \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

The trough indicator z_t^T is defined in an analogous way.[1] As we investigate monthly time series, we define $\tau = 2$. Choosing $\tau = 1$ would result in a model too sensitive to smaller movements of the time series, whereas with $\tau > 2$ the model would react with inacceptable delay. At time t the economist knows only the current and past datapoints $x_t, x_{t-1}, \ldots, x_{t-\tau+1}, x_{t-\tau}$. The future values $x_{t+1}, \ldots, x_{t+\tau-1}, x_{t+\tau}$ have to be estimated. Since the turning point indicators $z_t^P, z_{t+1}^P, \ldots,$ respectively $z_t^T, z_{t+1}^T, \ldots,$ are functions of the future datapoints $X_{t+1}, X_{t+2}, \ldots$, they are random variables. Using econometric models and the known $x_t, x_{t-1}, ..., x_{t-\tau+1}, x_{t-\tau}$, one can estimate $\hat{x}_{t+1}, \ldots, \hat{x}_{t+\tau-1}, \hat{x}_{t+\tau}$. Here it becomes clear that the ability for the detection of turning points critically depends on the forecasting model. The time series (X_t) could be generated by a univariate autoregressive process. In this case the following model is adequate to describe the true data generating process (DGP):[2]

$$x_{t+1} = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \ldots + \beta_R x_{t-R+1} + \epsilon_{t+1} \tag{3}$$

where β_i are the regression coefficients, R is the order of the AR process, and ϵ_{t+1} is a white noise disturbance term. Using optimisation techniques, such as Ordinary Least Squares, a model can be estimated from the data so that $E[\hat{\beta}_i] = \beta_i$. The model reflects the supposed DGP. The standard deviations $\sigma_{\hat{\beta}_i}$ of the estimated regression coefficients $\hat{\beta}_i$ and the standard deviation $\sigma_{\hat{\epsilon}_{t+1}}$ of the disturbance term are measures of the uncertainty of the forecast by the model and can be used to judge the ability of the model to mimic the true DGP. High $\sigma_{\hat{\beta}_i}$

[1] Furtheron we do not explicitly distinguish between peaks and troughs.

[2] The simple AR-process only serves for illustration purposes. More complex processes (e.g. VAR-, non-linear processes) could be relevant as well. The time series (X_t) here is meant to decribe a scalar. Generalisations to vector notation are straightforward.

resp. $\sigma_{\hat{\epsilon}_{t+1}}$ correspond to high uncertainty, whereas low values of $\sigma_{\hat{\beta}_i}$ resp. $\sigma_{\hat{\epsilon}_{t+1}}$ mean low uncertainty. After estimating $\hat{x}_{t+1}, \ldots, \hat{x}_{t+\tau}$ the indicators z_t^P and z_t^T can be computed. We are interested in a statement for the next turning point, so we define

$$w_t^P = w(z_{t-\tau+1}^P, z_{t-\tau+2}^P, \ldots) = k \tag{4}$$

where k is such that $z_{t+k}^P = 1$ and $z_{t+j}^P = 0, j < k$. Verbally interpreted w_t^P expresses the number of periods until the next turning point. The following 6-step Monte-Carlo procedure can be used to derive probabilistic statements for near-by turning points:

1. Draw random numbers $\tilde{\beta}_1(1), \tilde{\beta}_2(1), \ldots, \tilde{\beta}_R(1)$ from a multivariate normal distribution with mean vector[3] $(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_R)^T$ and empirical variance-covariance matrix of the regression coefficients.
2. Draw a random number $\tilde{\epsilon}_{t+1}(1)$ from a univariate normal distribution with mean 0 and variance $\sigma^2_{\hat{\epsilon}_{t+1}}$.[4]
3. Compute $\tilde{x}_{t+1}(1) = \tilde{\beta}_0(1)x_t + \tilde{\beta}_1(1)x_{t-1} + \ldots + \tilde{\beta}_R(1)x_{t-R+1} + \tilde{\epsilon}_{t+1}(1)$. If $\tau > 1$, draw $\tilde{\epsilon}_{t+2}(1), \ldots, \tilde{\epsilon}_{t+\tau}(1)$ and iterate step 3 to obtain $\tilde{x}_{t+2}(1), \ldots, \tilde{x}_{t+\tau}(1)$.
4. Compute $z_{t-\tau+1}^P(1), \ldots, z_{t+h-\tau}^P(1)$ and $z_{t-\tau+1}^T(1), \ldots, z_{t+h-\tau}^T(1)$.
5. Compute $w_t^P(1)$ and $w_t^T(1)$.
6. Repeat steps 1 to 5 N times.

The predictive distributions P_t^P and P_t^T for a near-by turning point can be approximated by the empirical distributions $w_t^P(1), \ldots, w_t^P(N)$ and $w_t^T(1), \ldots, w_t^T(N)$. As an example take the following table derived from a Monte-Carlo-Simulation with $N = 10$ drawings: There are three entries for a turning point in

Table 1. Example for determination of the turning point probabilities

n	1	2	3	4	5	6	7	8	9	10
$w_t^P(n)$	3	0	2	1	2	2	1	2	2	1

$t+1 : w_t^P(4) = 1, w_t^P(7) = 1, w_t^P(10) = 1$. It follows that $P_t^P(W_t^P = 1) = \frac{3}{10} = 0.3$. The probabilities for one period are characterised by $P_t^P + P_t^T \leq 1$. A turning point ist detected, if P_t^P reaches or exceeds a certain threshold θ, e.g. $\theta = 0.5$. Summarizing section 2, we explicitly incorporate uncertainty of the forecasts by producing probabilistic statements for near-by turning points. Furthermore, by not only considering one single model, but a family of N models, our forecasts are more reliable than those of single models. To compare the ARMA and VAR models, we need a measure of performance we base our decision on. It is discussed in the next section.

[3] The exponent T symbolizes transposition of the vector.

[4] Kling[2], p. 212, estimates the turning points with a VAR model. He draws the residual vector from a multivariate normal distribution with mean 0 and empirical variance-covariance matrix of the residuals.

3 Evaluation of Performance

The econometric model classes (ARMA and VAR) applied in this paper have to be evaluated concerning their task to forecast turning points. Hence it does not make sense in this context to rely on error measures of function approximation, such as MSE, MAE, etc. A participant in the financial markets usually is not interested in function approximation but in economic performance. Unfortunately error measures of function approximation show little coherence with trading profits[5]. Criteria especially developed to evaluate a model's ability to forecast turning points were developed, amongst others, by Brier[6] and Diebold/Rudebusch[7]. But those performance measures are similar to the error measures and a statistical test of significance is not available. One of our main goals in this study is evaluation based on economic criteria such as profits from a trading strategy. Since our models do not produce return forecasts but probabilities for turning points, we have to measure performance indirectly by generating trading signals from those probabilities: A short position is taken when a peak is detected ($P_t^P \geq \theta$, implying that the market will fall, trading signal s=-1), a long position in the case of a trough ($P_t^T \geq \theta$, market will rise, s=+1), and the position of the previous period is maintained if there is no turning point. One possibility to evaluate the quality of the trading signal forecasts is to count the number of correct forecasts and relate it to the number of incorrect ones. Unfortunately, this type of performance measurement does not discriminate between trading signals attributed to large and small movements of the time series. A participant in the financial markets usually is interested to get the large movements rather than the small ones. Correctly predicting the large movements corresponds to the idea of maximising a profit-oriented criterion. The Sharpe-Ratio, which is briefly described in the following, is such a criterion. With the actual period-to-period return $r_{actual,t}$ we can calculate the return $r_{m,t}$ from a turning point forecast of our model:

$$r_{m,t} = s \cdot r_{actual,t} \tag{5}$$

Subtracting the risk-free rate of interest $r_{f,T}$ from the average return of the model $\bar{r}_m = \frac{1}{T}\sum_{t=1}^{T} r_{m,t}$ over T periods yields the average excess return $\bar{e}_m = \bar{r}_m - r_{f,T}$. The Sharpe-Ratio SR[8] relates $\bar{e}_m$ to the return's standard deviation σ_m:

$$SR = \frac{\bar{e}_m}{\sigma_m} \tag{6}$$

The Sharpe-Ratio measures the excess return a model produces for a unit of risk the model takes. Jobson/Korkie[3] developed a test for the null hypothesis

$$H_0 : SR_m - SR_{BM} = 0 \tag{7}$$

of no significant difference between the Sharpe-Ratios of a forecasting model and a benchmark.[5] By re-arranging (7) and relating it to the variance δ of the two Sharpe-Ratios[6], Jobson/Korkie find that the test statistic

$$z_{m,BM} = \frac{\bar{e}_m \cdot \sigma_{BM} - \bar{e}_{BM} \cdot \sigma_m}{\sqrt{\delta}} \tag{8}$$

asymptotically follows a standard normal distribution and is powerful in moderately large samples. A positive and statistically significant $z_{m,BM}$ means that the model outperforms the benchmark in terms of the Sharpe-Ratio. The next section shows how the Jobson/Korkie-test was used within our Data Mining approach.

4 Data Mining in Financial Time Series

To test the ability of the ARMA and VAR models to predict turning points, we investigated the logarithms of nine financial time series, namely DMDOLLAR, YENDOLLAR, BD10Y (performance index for the 10 year German government benchmark bond), US10Y, JP10Y, MSWGR (performance index for the German stock market), MSUSA, MSJPA, and the CRB-Index. The data was available in monthly periodicity from 83.12 to 97.12, equalling 169 datapoints. One was lost because of differencing. 100 datapoints were used to estimate the models, so that we can base our decision which model class performs better on 68 out-of-sample forecasts. To allow for the possibility of structural change in the data, we implemented rolling regressions: After estimating the models with the first 100 datapoints and forecasting the succeeding datapoints, the data-"window" of the fixed size of 100 datapoints was put forth for one period and the estimation procedure was repeated. We estimated a multitude of models for each model class: 15 ARMA-models from (1,0), (0,1), (1,1),..., to (3,3) and 3 VAR models VAR(1), (2), and (3). We do not specify a model and estimate all rolling regressions with this model. Rather we specify a class of models (ARMA and VAR). Within a class the best model is selected for forecasting. As an extreme case, a different model specification could be chosen for every datapoint (within the ARMA class e.g. the ARMA(1,0) model for the first rolling regression, ARMA(2,2) for the second etc.). This model selection procedure is purely data-driven, so it can be regarded as Data Mining. Since it is well known that in-sample evaluation is a poor approximation for out-of-sample performance [10], reliable model selection has to be based on true out-of-sample validation. Therefore we divided the data in three subsequent, disjunct parts: a training subset (70 datapoints), a validation subset (30 datapoints), and a forecast subset ($\tau = 2$ datapoints, see Fig. 1).[7]

[5] Jobson/Korkie[3] developed a test statistic which allows to compare the Sharpe-Ratios of more than two portfolios simultaneously, too.

[6] $\delta = \frac{1}{T} 2\sigma_m^2 \sigma_{BM}^2 - 2\sigma_m \sigma_{BM} \sigma_{m,BM} + \frac{1}{2}\bar{e}_m^2 \sigma_{BM}^2 + \frac{1}{2}\bar{e}_{BM}^2 \sigma_m^2 - \frac{\bar{e}_m \bar{e}_{BM}}{2\sigma_m \sigma_{BM}}(\sigma_{m,BM}^2 + \sigma_m^2 \sigma_{BM}^2)]$.

[7] τ symbolizes the number of periods that have to be forecast in order to make a turning point decision at time t, cf. section 2.

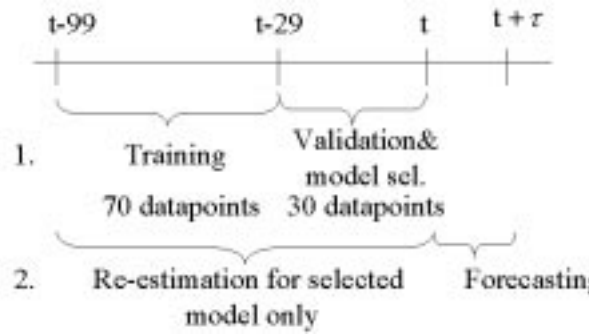

Fig. 1. Division of the database

The first 70 datapoints from t-99 to t-30 were used to estimate the models, which were validated with respect to their abilities to predict turning points on the following 30 datapoints from t-29 to t. This is true out-of-sample validation of the models, because at time t-30 the datapoints from t-29 to t are unknown. For each model and each datapoint in each rolling regression N=200 Monte-Carlo-simulations in order to calculate the turning point probabilities were performed. If the model at the beginning of the validation period in t-29 decided "no turning point" (results in maintenance of the previous period's trading signal), there is no trading signal originally stemming from the model. In this case we used the last trading signal which could be produced with certainty. With $\tau = 2$ the last certain signal for a turning point can be generated for t-31. Then the best model on the validation subset was selected. For each of the two model sequences (ARMA and VAR) only one model was selected at each time. The specification of this model, e.g. ARMA(2,2), then was re-estimated with the 100 datapoints from t-99 to t to forecast the at time t unknown τ values of the time series, which are necessary to decide whether there is a turning point at time t. The decision which model is the "best" was made with respect to the Jobson/Korkie-test on the difference between the Sharpe-Ratio of the models and a benchmark. Thus each of the multitude of models of each sequence had to be tested against a benchmark. Since our goal is the comparison of the two competing classes ARMA vs. VAR models, each sequence has to consist of representatives of this model class. The simplest model of each class served as a benchmark in the statistical tests: for the ARMA-sequence the benchmark was the (1,0)-model, in the other case the VAR(1)-model. Using the Jobson/Korkie-test as a criterion for selection, each of the multitude of ARMA models was tested against the (1,0) model. If e.g. the ARMA(2,2) model could reject the null hypothesis $z_{m,BM} = 0$ with a significantly positive $z_{m,BM}$ on the validation subset (t-29 to t), this model specification was selected and re-estimated with t-99 to t to forecast the τ unknown values of the time series for $t+1, ..., t+\tau$. If more than one model qualified, the one with the highest SR was selected. If the null could not be rejected, the forecasts were conducted with the (1,0)-model.

Table 2. ARMA- and VAR-sequence as an example for the rolling regressions

				ARMA		***VAR***	
RR	training	validation	forecast	Spec.	$z_{m,BM}$	Spec.	$z_{m,BM}$
1	84.1-89.10	89.11-92.4	92.5-92.6	(1,0)	*	(3)	3.3563 (.0008)
2	84.2-89.11	89.12-92.5	92.6-92.7	(2,2)	1.9823 (.0475)	(1)	*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
68	89.8-95.4	95.5-97.10	97.11-97.12	(3,0)	2.1486 (.0317)	(3)	1.9987 (.0456)

This procedure was implemented for the VAR-sequence in an analogous way. As we solely rely on economic performance to select the best model, we do not consider statistical criteria, like t-values etc. The variances resp. standard deviations of the regression coefficients are needed for the drawings of the random numbers to incorporate the uncertainty in the forecasts. Badly fitted models with high variability in the coefficients and according high variances will not be able to detect the relevant turning points in the validation subset and so be disqualified in the selection procedure. Two sequences with a threshold $\theta = 0.5$ and a significance level of 0.1 could look like Table 2. The first four columns refer to the number of the rolling regressions and the training, validation, and forecast period, respectively. The 5th (7th) column gives the specification of the selected ARMA (VAR) model, the 6th (8th) column gives $z_{m,BM}$ (the entry (*) in the column "$z_{m,BM}$" means that in this period no model qualifies against the benchmark; p-values in parenthesis below the $z_{m,BM}$-value):

The first turning point forecast was done for 92.4 (with the unknown values of 92.5 and 92.6), the last for 97.10. The primary objective of this paper is to make a statement about the relative performance of ARMA- vs. VAR-models to detect turning points in time series. We created two model sequences with a sample size of 68 forecasts each. In order to produce a statistically significant result, we compare the ARMA sequence with the VAR sequence. Therefore we compute the Sharpe-Ratios of the ARMA excess returns (SR_{ARMA}) and the VAR excess returns (SR_{VAR}) over the 68 out-of-sample rolling regression forecasts. Comparing those two Sharpe-Ratios with the Jobson/Korkie test statistic $z_{ARMA,VAR}$ (thereby using SR_{VAR} as the benchmark) allows to make a statement which model class performs better on the 68 datapoints. A result might be that $z_{ARMA,VAR} > 0$ and statistically significant (e.g. $z_{ARMA,VAR} = 1.96$ is significant at the 5%-level). In that case the conclusion is that ARMA models outperform VAR models.[8] This does not mean that ARMA models are valuable forecasting tools. In order to be sure that ARMA models in this example are valuable forecasting tools, one would like to test if this model class is able to

[8] In the case $z_{ARMA,VAR} < 0$ VAR-models outperform ARMA-models.

outperform a simple benchmark as well. When forecasting economic time series, a simple benchmark is the naive forecast. The naive forecast uses the status of the current period to forecast the next period. In the context of return forecasts, the return of the current period is extrapolated as a forecast of the next period. As we deal with turning point predictions and do not produce explicit return but trading signal forecasts derived from turning point predictions, the extrapolation of returns is not adequate. One could think of using the actual return from t-1 to t as an indicator for the trading signal of the future period: a past positive return means a trading signal +1 for the future period, a negative return means a trading signal -1. One goal of turning point predictions is to detect the longer term trend reversals. Using the past return as a benchmark is more adequate for short-term, period-to-period forecasts. Hence we need a naive benchmark which works in a similar way as our model and thereby reflects the idea of turning point forecasts.[9] Using the last certain turning point statement can be regarded as a benchmark in this sense. As $\tau = 2$, the last certain turning point statement can be made for $t - 2$, using the datapoints from $t - 4$ to t. A valuable forecasting model should be able to outperform this Naive Turning Point Forecast (NTPF), so it is straightforward to test the ARMA- and VAR-sequences against the NTPF. E.g. a significantly positive $z_{ARMA,NTPF}$ ($z_{VAR,NTPF}$) calculated with the Sharpe-Ratios for the ARMA- (VAR-) and NTPF-sequences over the 68 out-of-sample forecasts implies that e.g. the ARMA- (VAR-) models are valuable tools for forecasting turning points in financial time series. If $z_{ARMA,NTPF}$ is negative, the NTPF produces a higher Sharpe-Ratio over the 68 datapoints. The next section presents empirical results from the turning point forecasts with our Data Mining approach.

5 Empirical Results

The following Table 3 exhibits the results for the turning point forecasts with a significance level of SL=.1 for the Jobson/Korkie-test and a threshold of $\theta = .75$. Results with different threshold values $\theta = .5$ and $\theta = .95$ showed that overall $\theta = .75$ produces the best results of the model classes vs. NTPF. The lower the level of significance for the Jobson/Korkie-test, the higher is the required difference between the two Sharpe-Ratios to be considered as statistically significant. With SL=.01, the sequences almost solely comprised the simplest (=benchmark) model (ARMA(1,0) resp. VAR(1)) in each model sequence. To a lesser degree this was valid for the turning point forecasts with a significance level SL=.1, whose results are presented in Table 3. Another point in this direction comes from the sample size of only 30 datapoints in the validation subset, which results in low discriminative power of the test. Hence the bias for the selection of the simplest model in each sequence is relatively high. The nine financial time series are considered as a closed market system, where every variable influences each other. Therefore the VAR models consisted of nine equations with lags of all

[9] For a detailed discussion of the problem to select an adequate benchmark for turning point models see Poddig/Huber[11], p. 30ff.

Table 3. Empirical results

				ARMAvsVAR		*ARMAvsNTPF*		*VARvsNTPF*	
$\theta = .75$	SR_{ARMA}	SR_{VAR}	SR_{NTPF}	z	$p(z)$	z	$p(z)$	z	$p(z)$
MSWG	.0124	-.1419	-.0038	1.44	.1488	.13	.8897	-.94	.3450
MSUSA	.1149	-.0287	.0341	1.10	.2703	.71	.4774	-.49	.6265
MSJPA	-.0398	.1300	-.0700	-1.14	.2544	.13	.8955	1.07	.2862
BD10Y	-.2206	-.3545	-.2602	1.39	.1653	.47	.6402	-.78	.4352
US10Y	-.1253	-.0851	-.1597	-.36	.7187	.27	.7855	.50	.6178
JP10Y	.0883	.0480	-.0824	.58	.5613	1.35	.1764	1.16	.2450
DMDO	-.2569	-.3169	-.3098	.45	.6520	.39	.6977	-.05	.9636
YEND	-.1966	-.1931	.1015	-.03	.9755	-.52	.5992	-.56	.5778
CRB	-.3458	-.4307	-.4759	.44	.6597	1.05	.2920	.24	.8132

variables in the system and a constant. Table 3 exhibits in the first column the name of the time series under consideration. The three following columns show the values of the Sharpe-Ratios of the ARMA- and VAR-models, and the NTPF, respectively. For each change from a long- into a short-position transaction costs of 0.75% were subtracted. The 5th column and 6th column give the z-values for the Jobson/Korkie-test statistic and its corresponding p-value for the test of the ARMA- vs. the VAR-sequences. The two following columns present z- and p-values for the comparison of the ARMA- vs. NTPF-sequences. The 9th and 10th column exhibit those values for the VAR- vs. NTPF-sequences.

Looking at the results for e.g. MSWGR in detail, only the ARMA-models were able to produce a positive Sharpe-Ratio ($SR_{ARMA} = .0124$) in the out-of-sample forecasts for the 68 months under consideration. This means that the ARMA-models on average reward each unit of risk, measured in standard deviations of the return, with an excess return of 1.24% per month. From the positive $z_{ARMA,VAR} = 1.44$ ($p = .1488$) for the nullhypothesis $H_0 : SR_{ARMA} - SR_{VAR} = 0$ it can be seen that they outperformed VAR models, although not significantly at the usual levels (1% to 10%). ARMA-models outperformed the NTPF as well ($z_{ARMA,NTPF} = .13, p = .8897$). VAR-models underperformed even the NTPF, so they do not seem to be a valuable tool for forecasting turning points in the German stock market. ARMA-models managed to produce positive Sharpe-Ratios for three out of the nine markets (MSWGR, MSUSA, JP10Y), VAR (MSJPA, JP10Y) and NTPF only twice (MSUSA, YENDOLLAR). In six markets ARMA-models outperformed VAR-models (positive $z_{ARMA,VAR}$-values), and in all markets but YENDOLLAR they outperformed the NTPF (positive $z_{ARMA,NTPF}$-values), but not significantly. VAR-models manage to beat the NTPF only four times (positive $z_{VAR,NTPF}$–values for MSJPA, US10Y, JP10Y, CRB). MSJPA is the only market in which VAR-models remarkably outperform ARMA-models ($z_{ARMA,VAR} = -1.14, p = .2544$). In two more cases VAR at least produce higher Sharpe-Ratios ($z_{ARMA,VAR}$ negative for US10Y and YENDOLLAR). Only the Japanese stock market MSJPA could be predicted remark-

ably more successfully by VAR- than by ARMA-models and NTPF (MSJPA: $z_{VAR,NTPF} = 1.07; z_{ARMA,VAR} = -1.14$. For JP10Y the VAR outperformed the NTPF, but so did the ARMA-models. The CRB-index seems to be unpredictable: none of the three sequences produced a positive Sharpe-Ratio, but the ARMA-(SR=-.3458) and VAR-models (SR=-.4307) still performed better than the NTPF (SR=-.4759). Summarizing the results in brief, it seems that ARMA-models are better tools for forecasting turning points in financial time series. In all but one case they managed to outperform the NTPF, although not in one single case statistically significant. The bad performance of the VAR-models might be due to their possible overparameterisation with nine variables and one to three lags. The simplest VAR(1)-model comprises nine variables plus a constant in each equation, which results in (9+1)·9=90 regression coefficients. Future research in the area of turning point forecasts will concentrate on smaller VAR-models and the derivation of portfolio weights from the turning point probabilities. This can be accomplished by "rewarding" forecasts with a high degree of certainty in the forecasts (e.g. $P_t^P \approx 1 \Rightarrow$ certainty of a peak, or $P_t^P \approx 0 \Rightarrow$ certainty of no peak) and "penalizing" the ones with a low degree of certainty ($P_t^P \approx \theta$).

References

1. Wecker, W. (1979): Predicting the turning points of a time series; in: Journal of Business, Vol. 52, No. 1, 35-50
2. Kling, J.L. (1987): Predicting the turning points of business and economic time series; in: Journal of Business, Vol. 60, No. 2, 201-238
3. Jobson, J.D.; Korkie, B.M. (1981): Performance Hypothesis Testing with the Sharpe and Treynor Measures, in: Journal of Finance, Vol. XXXVI, No. 4, 889-908
4. McLeod, G. (1983): Box-Jenkins in Practice, p. 11-130, Lancaster
5. Leitch, G.; Tanner, E. (1991): Economic Forecast Evaluation: Profits versus the conventional error measures, p. 584; in: American Economic Review, Vol. 81, No.3, 580-590
6. Brier, G.W. (1950): Verification of forecasts expressed in terms of probability; in: Monthly Weather Review, Vol. 75 (January), 1-3
7. Diebold, F.X.; Rudebusch, G.D. (1989): Scoring the Leading Indicators; in: Journal of Business, Vol. 62, No. 3, 369-391
8. Sharpe, W.F. (1966): Mutual Fund Performance; in: Journal of Business, 119-138
9. Swanson, N.R.; White, H. (1997): Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models; in: International Journal of Forecasting, Vol. 13, 439-461
10. Poddig, Th. (1996): Analyse und Prognose von Finanzmärkten, Bad Soden/Ts., p. 380ff.
11. Poddig, Th.; Huber, C. (1998): Data Mining mit ARIMA-Modellen zur Prognose von Wendepunkten in Finanzmarktzeitreihen; Discussion Papers in Finance, No. 1, University of Bremen, available at http://www1.uni-bremen.de/~fiwi/