GERHARD THURY

Forecasting Performance of Structural Time Series Models

A Case Study for Austrian and German Industrial Production

64

GERHARD THURY

Forecasting Performance of Structural Time Series Models

A Case Study for Austrian and German Industrial Production

October 1993

WIFO Working Papers, 1993, (64)

Gerhard Thury

Forecasting Performance of Structural Time Series Models

A Case Study for Austrian and German Industrial Production

In a recent paper, Hahn and Thury (1992) present estimates of structural time series models for the Austrian and German industrial production. A comparison of the forecasting performance of this model type with that of traditional univariate time series models might be informative. As measure of reference, we use the multiplicative seasonal *ARIMA* model, which is widely used in applied economic forecasting. We have 372 monthly observations on the indices of industrial production in Austria and Germany. Since calendar variations strongly influence production data, the two series are first adjusted for these effects before the different time series models are estimated. Details about this calendar adjustment can be found in Thury (1989). Since the number of observations at our disposal is relatively large, we can retain a substantial portion of these observations in order to test the forecasting performance of the estimated time series models. Thus, for the different forecasting methods and forecasting horizons, we generate 120 genuine ex-ante predictions covering the period 1983:1 to 1992:12 which, then, form the basis for an evaluation of the forecasting accuracy of the methods under test. We are convinced that 120 observations should be sufficient to derive reliable estimates for various test statistics of forecasting performance.

The organisation of the paper is as follows. We start out with a short description of the measures of forecasting accuracy which we shall employ in this paper. The main part of the paper consists of a presentation and interpretation of our empirical results. In a short concluding section, finally, we summarize our main findings.

Theoretical considerations

In assessing the forecasting accuracy of the time series models under consideration, we closely follow the path proposed by Witt and Witt (1992). We begin with a detailed analysis of the committed forecast errors. Since it is often claimed, however, that directional accuracy is, at least, as important as the magnitude of the forecast error, we investigate the performance in this respect also very carefully.

Measures of numerical accuracy and statistical tests

In order to evaluate the accuracy of a forecasting method, it is necessary to have a yardstick. There exist various measures of forecasting accuracy but, unfortunately, none of them is universally accepted. Following Witt and Witt, we shall concentrate on two relative measures of forecasting accuracy, namely the mean absolute percentage error and the root mean square percentage error.

The mean absolute percentage error (*MAPE*) is given by

$$(1) \qquad MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|e_t|}{IP_t} \cdot 100,$$

where $|e_t|$ denotes the absolute value of the forecast error and n is the number of forecasts. The forecast error is given by

$$(2) \qquad e_t = \hat{IP}_t - IP_t ,$$

where $\hat{IP}_t$ and IP_t symbolize predicted and actual values of the index of industrial production, respectively. *MAPE* is a measure of overall accuracy which offers an indication of the degree of spread between predicted and observed values. All forecast errors are assigned equal weights. Table 1 contains typical *MAPE* values for industrial data and their interpretation, which were published originally by Lewis (1982).

Table 1

Interpretation of typical *MAPE* values

MAPE	Interpretation
< 10 precent	Highly accurate forecasting
10 – 20 percent	Good forecasting
20 – 50 percent	Reasonable forecasting
> 50 percent	Inaccurate forecasting

Reproduced from Witt and Witt(1992), p. 86

The root mean square percentage error (*RMSPE*) is given by

$$(3) \qquad RMSPE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left[\frac{e_t}{IP_t} \right]^2} \cdot 100.$$

The *RMSPE* is also a measure of overall accuracy which provides an indication of the degree of spread. But, contrary to *MAPE*, large errors are penalized by additional weight.

In studies of forecasting performance it has been common practice for a long time to simply present accuracy measures in tabular form. Supplementing these presentations by statistical tests will provide additional insight and might allow firmer conclusions. An approach, which seems to be especially adequate for the purposes of this paper, is the ANOVA technique because it allows for varying numbers of factors to be tested simultaneously. Thus, we can test whether there exist significant differences between forecasting methods, forecasting horizons, and production countries. A certain drawback of the ANOVA approach lies in the fact that it may indicate significant differences among factor levels, but not between which levels if there are more than two. Multiple comparison tests, as for example Scheffe's test, or pairwise t-tests can provide answers to open questions of this type.

Measures of directional accuracy

Numerical accuracy is one of the desirable features of a forecast, directional accuary is another, perhaps even more important, property. With directional accuary, we must distinguish between direction of change errors (sometimes also called tracking errors) and trend change errors. A direction of change error occurs if the forecast misses the actual direction of change. There are several possibilities. The predicted change is positive and the actual change is negative or vice versa. Additionally, an observed change in direction can be missed by the forecast or a change in direction can be predicted which, then, does not realize. We compress these different possibilities into a single measure of direction of change error by calculating the percentage of correctly predicted changes of direction.

Instead of just looking generally at directional accuracy, it may be informative to analyze the situation more closely by examining trend change accuracy. A trend change error is observed when either a forecasting method fails to predict a realized change in the trend (a missed trend change) or incorrectly predicts a trend change (a false signal). Trend changes may be divided into downturns and upturns, and varying numbers of observations can be employed in their definition. Following Witt and Witt, we define them as follows:

$$(4) \quad y_{n-2} < y_{n-1} < y_n \quad \text{and} \quad \begin{cases} Z < y_n & = \text{Downturn (DT)}, \\ Z \geq y_n & = \text{No downturn (NDT)}, \end{cases}$$

and

$$y_{n-2} > y_{n-1} > y_n \quad \text{and} \quad \begin{cases} Z > y_n & = \text{Upturn (UT)}, \\ Z \leq y_n & = \text{No upturn (NUT)}, \end{cases}$$

where y_1 , y_2 , . . ., y_n denote given past realizations of a time series, and $Z \equiv y_{n+1}$ is the first future value of this series. Four consecutive observations are used to define downturns and upturns. A downturn is observed when an increasing trend has been established by the two observations preceeding the current one and the following observation is smaller than the current one. Similarly, an upturn occurs when a decreasing trend has been established, and the following observation is greater than the current one.

Empirical results

In the following, we analyze the forecasting performance of four univariate time series models:

basic structural model (*BSM*);

structural model with additive cycle (*SMAC*);

structural model with additive cycle and damping factor 1.00 (*SMACX*);

Box-Jenkins airline model (*ARIMA*).

These four model versions are used to generate predictions for the Austrian and German industrial production with forecasting horizons of 1, 6, 12, 18, and 24 months covering the period 1983:1 to 1992:12. In order to obtain these forecasts, the models are always reestimated for the relevant sample periods. Only information, which would have been available already at the date of the forecast origin, is utilized. The predictions, which are analyzed in this paper, are thus genuine ex-ante forecasts[1]).

Numerical measures of forecasting accuracy

Tables 2 and 3 contain measures of forecasting accuarcy for the different sets of predictions. Table 2 summarizes the results for *RMSPE*, Table 3 those for *MAPE*. Generally, both *RMSPE* and *MAPE* are subject to distortion caused by outlying observations, in that one or two poor forecasts will affect these average error measures. However, given the large number of forecast errors under analysis, this is unlikely to be a serious problem here.

Before entering into a detailed discussion of the different accuracy measures, some form of general assessment might be useful. Comparing the *MAPE*'s in Table 3 with the typical values given in Table 1, we see that, according to this standard, even the 24-months-ahead predictions figure as highly

---

[1]) In the computations for this paper the PC versions of the following programs are applied: SCA, SPSS, STAMP, and TSP. I wish to thank Sonja Patsios for her short, but very informative introductory course to SPSS.

accurate. Our judgement is less euphoric. We would say that forecasts with a horizon of up to 12 months might provide valuable information.

Table 2

Forecasting performance by forecasting horizon, forecasting method, and production country: *RMSPE*

Forecasting horizon (months)	Forecasting method	Production country: Austria	Production country: Germany
1	*BSM*	1.644(1)	1.734(1)
	SMAC	1.657(2)	1.741(2)
	SMACX	1.674(3)	1.765(4)
	ARIMA	1.695(4)	1.744(3)
6	*BSM*	2.675(1)	2.605(1)
	SMAC	2.757(2)	2.870(2)
	SMACX	2.856(4)	3.157(4)
	ARIMA	2.761(3)	2.933(3)
12	*BSM*	3.626(1)	3.002(1)
	SMAC	3.658(2)	3.678(2)
	SMACX	3.901(3)	4.279(4)
	ARIMA	4.039(4)	3.750(3)
18	*BSM*	4.793(1)	3.934(1)
	SMAC	4.891(2)	4.943(2)
	SMACX	5.157(3)	6.006(4)
	ARIMA	5.613(4)	5.090(3)
24	*BSM*	5.593(2)	4.377(1)
	SMAC	5.589(1)	5.943(3)
	SMACX	5.818(3)	7.421(4)
	ARIMA	6.714(4)	5.888(2)

Now, we shall turn to a detailed analysis of these accuracy measures. As expected has the length of the forecasting horizon the biggest effect for accuracy. Both measures, *RMSPE* and *MAPE*, give identical results in this respect. The longer the forecasting horizons the larger are the errors, although the differences for longer horizons (18 and 24 months) are less pronounced. The different forecasting methods also give rise to variations in the size of the forecast errors. The observed differences are however by far less significant than in the case of forecasting horizons. Additionally, we observe here slight discrepancies between the results for Austria and Germany. For Germany both measures, *RMSPE* and *MAPE*, yield identical results. Here, the ranking of forecasting methods is uneqivocal. *BSM* is best, followed by *SMAC*. It is perhaps a little surprising that *ARIMA* outperforms *SMACX*. For Austria, the outcome is slightly more controversial, as we observe certain differences in the ranking according to *RMSPE* and *MAPE*. Relying on *RMSPE*'s, the ranking is identical to the German one apart from the fact that, in Austria, the *ARIMA* model gives the worst forecasts. The *MAPE* based results also show that the structural time series models clearly outperform the traditional *ARIMA* model. Only the ranking within the group of structural models varies somewhat between the *RMSPE* and the *MAPE* results.

Table 3

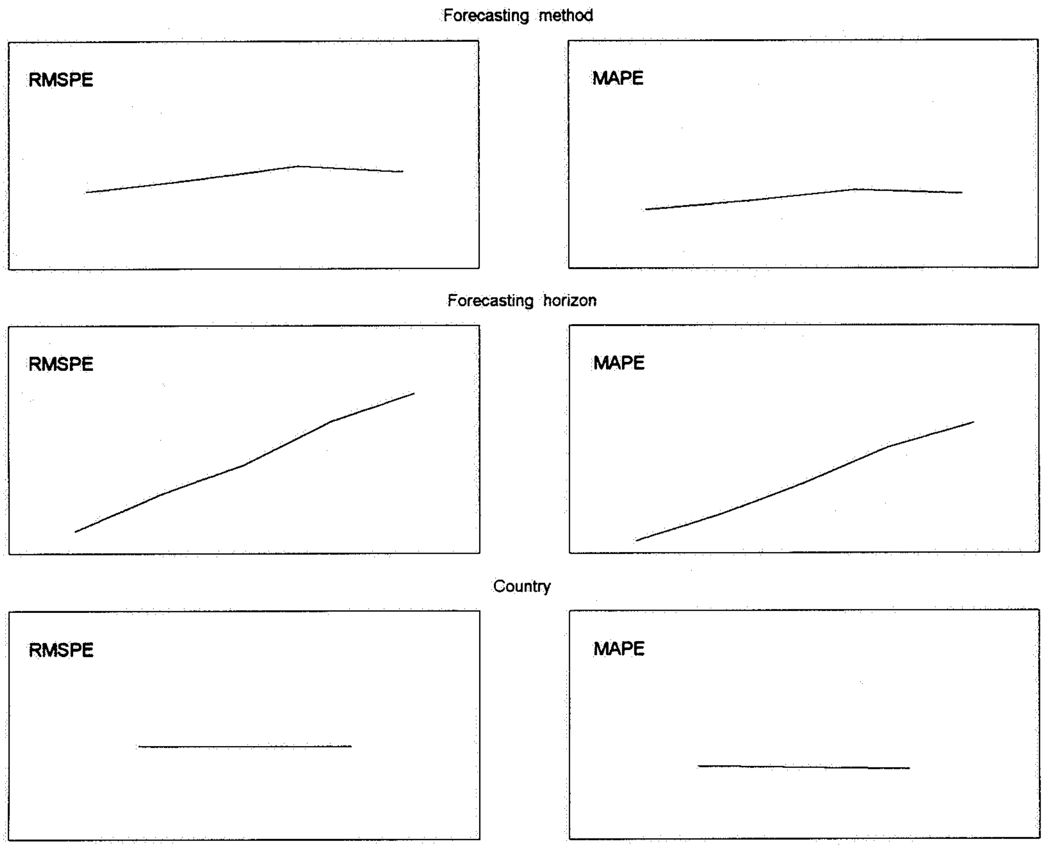
Forecasting performance by forecasting horizon, forecasting method, and production country: *MAPE*

Forecasting horizon (months)	Forecasting method	Production country: Austria	Production country: Germany
1	*BSM*	1.355(1)	1.382(2)
	SMAC	1.373(3)	1.378(1)
	SMACX	1.372(2)	1.389(4)
	ARIMA	1.393(4)	1.388(3)
6	*BSM*	2.125(3)	2.063(1)
	SMAC	2.053(1)	2.226(2)
	SMACX	2.119(2)	2.511(4)
	ARIMA	2.221(4)	2.228(3)
12	*BSM*	3.192(3)	2.498(1)
	SMAC	3.026(1)	3.069(2)
	SMACX	3.138(2)	3.556(4)
	ARIMA	3.592(4)	3.176(3)
18	*BSM*	4.127(1)	3.245(1)
	SMAC	4.179(2)	4.143(2)
	SMACX	4.286(3)	4.919(4)
	ARIMA	4.905(4)	4.306(3)
24	*BSM*	4.794(1)	3.583(1)
	SMAC	4.853(2)	4.926(3)
	SMACX	4.996(3)	6.217(4)
	ARIMA	5.782(4)	4.835(2)

Summarizing the outcome of this ranking exercise, one might say that, in both countries, a basic structural time series model gives the predictions of industrial production with the smallest forecast error. This result is somewhat surprising. Intuitively, one would expect that the explicit introduction of a cyclical component should improve the tracking performance of a model. But, although estimates of such cyclical components may provide valuable qualitative information on the current state of the business cycle, their extrapolation into the future seems to be too unreliable in order to give rise to improved quantitative forecasts. This problem becomes particularly visible, if the cyclical component enters a model in undamped form, as it is the case in *SMACX*.

Figure 1

Main effects for different factors

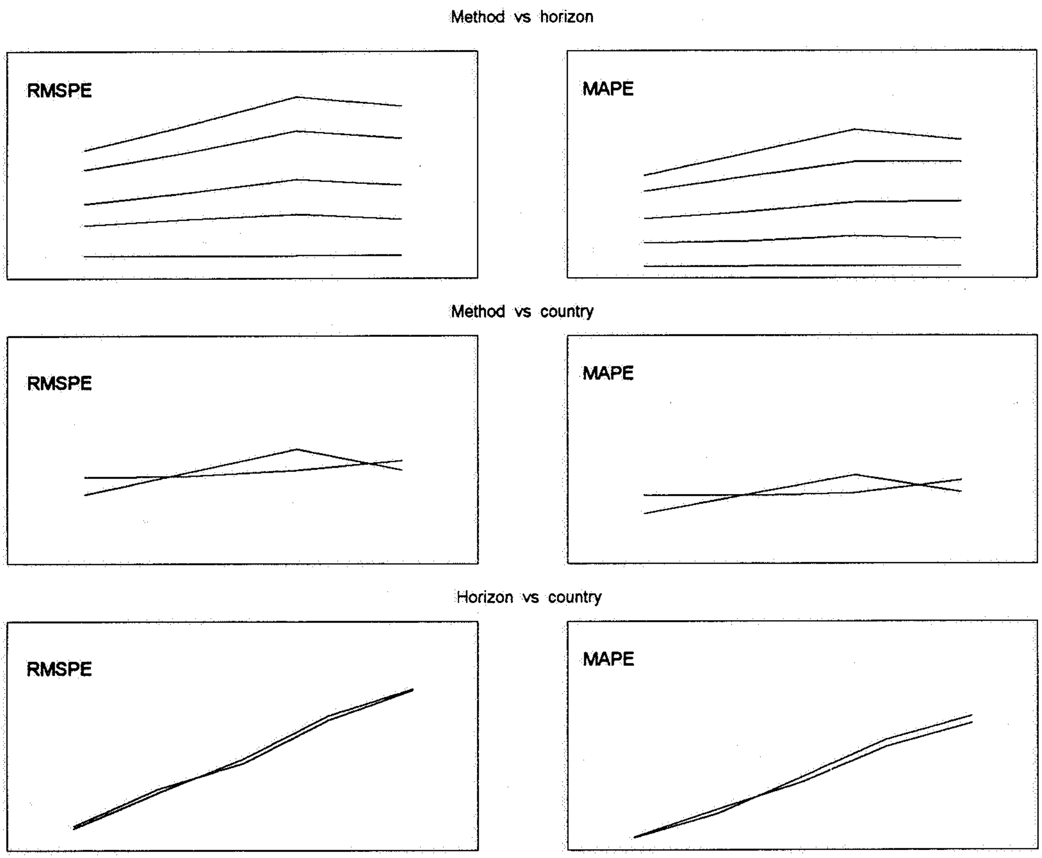


ANOVA results and statistical testing

Many empirical studies of forecasting performance present the results in terms of the accuracy measures considered. Often however additional analysis, such as an application of the ANOVA technique, might provide valuable further insight. There exist simple graphical techniques, which allow to decide quickly whether such further analysis is worthwhile or not. Figures 1 and 2 contain such simple graphs. In Figure 1 we depict the main effects for different factors individually. Thus, we have only one line in a graph. If this line slopes, this is an indication for the existence of a significant effect of a particular factor. The steeper the slope, the more significant the effect will be. If the line is horizontal, no significant effect is present. In Figure 2 we depict interaction effects for pairs of factors. Consequently, we have two lines in a graph. If these lines are parallel, no interaction between the factors occurs. If these lines intersect, significant interaction between two factors is present. Cursory inspection of these graphs provides some evidence for the existence of both main and interaction effects. Thus, further analysis seems to be worthwhile.

Figure 2

Interaction effects for different factors



Tables 4 and 5 present the outcome of an application of the ANOVA technique to *RMSPE*'s and *MAPE*'s. In each ANOVA table, we investigate if any of the three factors (forecasting method *FM*, forecasting horizon *H*, production country *C*) or interactions of these three factors have a significant effect on forecasting accuracy, as measured either by *RMSPE* or *MAPE*. *FM* has four levels, *H* five, and *C* two. Contrary to the above ranking example, the ANOVA results for *RMSPE*'s and *MAPE*'s are fully identical. In both cases, we find statistically significant effects for two of the three factors, namely for forecasting methods (*FM*) and forecasting horizon (*H*). The production country has no effect on the accuracy of the forecasts. There does not exist a statistically significant difference in the size of forecast errors for Austria and Germany. Of the three possible two-way interactions only one is significant, namely that between forecasting methods and production country. This implies that there exists a significant difference in the performance of the different forecasting methods in the two countries under study. We did not have enough data to test for the existence of a three-way interaction.

Table 4

ANOVA results for pooled data of Austria and Germany: ***RMSPE***

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	Significance of *F*
Main effects	94.661	8	11.833	89.704	0.000
Forecasting methods	3.647	3	1.216	9.217	0.002
Forecasting horizon	91.012	4	22.753	172.493	0.000
Production country	0.002	1	0.002	0.012	0.915
Two-way interactions	4.046	19	0.213	1.614	0.199
$FM-H$	1.945	12	0.162	1.229	0.363
$FM-C$	1.994	3	0.665	5.039	0.017
$H-C$	0.107	4	0.027	0.203	0.932
Explained	98.707	27	3.656	27.715	0.000
Residual	1.593	12	0.132		
Total	100.290	39	2.572		

Table 5

ANOVA results for pooled data of Austria and Germany: ***MAPE***

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	Significance of *F*
Main effects	72.008	8	9.001	97.171	0.000
Forecasting method	2.342	3	0.781	8.427	0.003
Forecasting horizon	69.582	4	17.395	187.792	0.000
Production country	0.085	1	0.085	0.917	0.357
Two-way interactions	3.542	19	0.186	2.012	0.108
$FM-H$	1.529	12	0.127	1.376	0.295
$FM-C$	1.821	3	0.607	6.553	0.007
$H-C$	0.191	4	0.048	0.517	0.725
Explained	75.550	27	2.798	30.207	0.000
Residual	1.112	12	0.093		
Total	76.661	39	1.966		

The above ANOVA results tell us only that there exist significant differences in forecasting accuracy for different forecasting horizons and different forecasting methods. However, they provide no ranking for different levels of a particular factor. Thus, if a significant factor has more than two levels, further statistical tests become necessary in order to find out which of these levels are different from each other.

Table 6

Modified pairwise *t*-tests for differences in accuracy between forecasting methods: pooled sample

RMSPE

	BSM	*SMAC*	*SMACX*	*ARIMA*
BSM	.			
SMAC	3.744**	.		
SMACX	8.051**	4.307**	.	
ARIMA	6.244**	2.500	1.807	.

MAPE

	BSM	*SMAC*	*SMACX*	*ARIMA*
BSM	.			
SMAC	2.862	.		
SMACX	6.139**	3.277*	.	
ARIMA	5.462**	2.600	0.677	.

* indicates 5% level of significance,
** indicates 1% level of significance

Of our two significant factors, forecasting horizon has five and forecasting method four levels. In order to determine a hierarchy among these levels, we apply various alternatives of multiple range tests (among them Scheffe's test). Only for forecasting horizons significant differences are revealed by these tests. *MAPE*'s and *RMSPE*'s are different for each particular forecasting horizon. That forecasts with shorter horizons have smaller errors, is selfevident. That all five horizons are classified as significantly different by multiple range tests, is somewhat surprising. For forecasting methods on the other side, no significant differences could be found by these tests. The reason for this failure may be twofold. Multiple range tests are, in general, rather conservative and the observed differences in the *MAPE*'s and *RMSPE*'s are numerically small. We use a modified version of a paired *t*-test in order to overcome these problems. The null of nonsignificant differences between forecasting methods is rejected if

$$(5) \quad | x_{i.} - x_{j.} | > \sqrt{2\, n\, S_R\, F_{n-a}}\,.$$

In this expression, $x_{i.}$ and $x_{j.}$ denote sample sums of *MAPE*'s and *RMSPE*'s, respectively, for method i and method j, n is the sample size, a the number of levels, and SR the residual mean square of the

ANOVA table with ($n-a$) degrees of freedom. The outcome of this testing procedure is given in Table 6. We see that for both accuracy measures, *MAPE* and *RMSPE*, *BSM* is definitely superior to *SMACX* and *ARIMA*. There exists some evidence that it also dominates *SMAC*. The poor performance of *SMACX* is an obvious consequence of setting the damping factor equal to 1.00. This operation is far from optimal in the context of forecasting. It might, however, provide valuable qualitative information on the future cyclical development of industrial production.

Of the three possible interaction effects only that between forecasting methods (*FM*) and production country (*C*) is statistically significant. There exist apparently some differences between Austria and Germany in the accuracy of the tested forecasting method. The situation can be looked at from two perspectives:

(i) the factor *FM* is specified and the data, on which the analysis is performed, are restricted to a given country;

(ii) the factor *C* is specified, and the data are restricted to a given method.

The first approach shows where any statistical differences lie among forecasting methods for a given country. The second approach informs us about differences between the two countries for a given forecasting method.

Table 7

Modified pairwise *t*-tests for interaction effects between forecasting methods and production country: *RMSPE*

Case (i)

	Austria				Germany			
	BSM	*SMAC*	*SMACX*	*ARIMA*	*BSM*	*SMAC*	*SMACX*	*ARIMA*
BSM	.				.			
SMAC	0.221	.			3.523*	.		
SMACX	1.075	0.854	.		6.976**	3.453*	.	
ARIMA	2.551**	2.330*	1.476	.	3.753*	0.230	3.223*	.

Case (ii)

		Austria			
		BSM	*SMAC*	*SMACX*	*ARIMA*
	BSM	2.679*			
Germany	*SMAC*		0.623		
	SMACX			3.222*	
	ARIMA				1.417

* indicates 5% level of significance,
** indicates 1% level of significance

Table 8

Modified pairwise *t*-tests for interaction effects between forecasting methods and production country: *MAPE*

Case (i)

	Austria				Germany			
	BSM	*SMAC*	*SMACX*	*ARIMA*	*BSM*	*SMAC*	*SMACX*	*ARIMA*
BSM	.				.			
SMAC	0.109	.			2.971*	.		
SMACX	1.318	0.427	.		5.821**	2.850	.	
ARIMA	2.300*	2.409*	1.982	.	3.162*	0.191	2.659	.

Case (ii)

		Austria			
		BSM	*SMAC*	*SMACX*	*ARIMA*
Germany	*BSM*	2.822*			
	SMAC		0.258		
	SMACX			2.681*	
	ARIMA				1.960

* indicates 5% level of significance,
** indicates 1% level of significance

Since multiple range tests yield no significant results, we apply the above mentioned modified *t*-tests. The outcome of these tests is presented in Tables 7 and 8. For both measures of accuracy, *RMSPE* and *MAPE*, we observe differences in the performance of the forecasting methods for a given country. These observed differences are for Germany much more significant than for Austria. For Germany, we find that *BSM* definitely outperforms the other methods. *SMAC* and *ARIMA* yield predictions of similar absolute accuracy. *SMACX* gives by far the worst results. For Austria, we find only weak evidence that *BSM* and *SMAC* do somewhat better than the other two methods, at least, as far as absolute forecasting accuracy is concerned. When looking for differences between the two countries for a given method, we find that *BSM* does significantly better in Germany than in Austria while, for *SMACX*, the exact opposite is the case. The performance of the remaining two methods is similar in both countries.

Directional forecasting accuracy

It is often argued by practitioners that, in situations where both goals cannot be achieved simultaneously, a forecasting method with a high degree of directional accuracy is preferable to a method with better absolute accuracy.

Directional change accuracy

Naive no-change extrapolations are neither correct nor incorrect predictions of the direction of change in a time series. A forecasting method outperforms such a naive model, if it can predict over 50% of the directions of change which occur. With monthly, seasonally unadjusted data, we have two possibilities to measure the directions of change: first differeneces (against the previous month) and annual differences (against the corresponding month of the previous year). The month-to-month differences of industrial production, for which strong seasonal movements are characteristic, exhibit wild fluctuations but are stationary in general. Annual differences, on the other side, are far less volatile but, often, a significant time trend is present. Moreover, by forming annual differences much information is lost. Since the results for month-to-month and annual differences diverge substantially for the two production series under analysis, we decided to present both. They are given in Tables 9 and 10. What we present in these tables are the percentages of directions of change predicted correctly by a particular forecasting method.

Table 9

Forecasting performance: Direction of change error in month-to-month differences

Percentage of direction changes forecast correctly

Forecasting methods	Forecasting horizon in months 1	6	12	18	24
Austria					
BSN	88.2	92.4	91.6	89.9	91.6
SMAC	88.2	89.9	91.6	89.1	91.6
SMACX	89.9	87.4	86.6	88.2	89.9
ARIMA	89.1	91.6	90.8	89.1	89.9
Germany					
BSM	95.8	95.0	95.8	94.1	96.6
SMAC	95.8	92.4	93.3	90.8	91.6
SMACX	95.0	90.8	87.4	83.2	79.0
ARIMA	95.8	95.0	93.3	89.9	95.0

We begin with a discussion of the results for first differences. We are somewhat surprised that here the relative forecasting performance is practically independent of the length of the forecasting horizon. The percentage of directions of change, which are predicted correctly by a particular method, hardly decreases for longer horizons (*SMACX* in Germany is the only exception). The explanation lies in the fact that we have eliminated the trend by taking first differences. Erratic fluctuations in the slope of a trend line seem to be the main cause of big forecast errors. The rest, that remains after removing the trend, is dominated strongly by seasonal fluctuations and, obviously, can be predicted very accurately. A substantial difference in the forecasting performance between Austria and

Germany is also worth noting. Apart from *SMACX*, all other forecasting methods exhibit a better relative performance in Germany, whereby the high percentages of correctly forecast month-to-month changes for *BSM* and, most surprisingly, *ARIMA* are especially remarkable. One possible explanation might lie in more stable climatic conditions in Germany.

Table 10

Forecasting performance: Direction of change error in annual differences

Percentage of direction changes forecast correctly

Forecasting methods	Forecasting horizon in months				
	1	6	12	18	24
Austria					
BSN	91.7	81.5	75.0	65.7	62.0
SMAC	91.7	81.5	75.0	66.7	64.8
SMACX	91.7	82.4	69.4	64.5	68.5
ARIMA	90.7	81.5	71.3	61.1	54.6
Germany					
BSM	82.4	81.5	72.2	63.0	58.3
SMAC	83.3	83.3	72.2	58 3	54.6
SMACX	85.2	82.4	67.6	52.8	49.1
ARIMA	83.3	79.6	69.4	58.3	52.8

Removing the seasonality by annual differencing and leaving the trend untouched yields results, which are more conform to expectations. In both countries, the relative performance deteriorates continuously with the increasing length of the forecasting horizon. This substantial decrease in the percentage of correctly predicted directional changes is caused by errors in the trend extrapolations.

The completely different relative forecasting performance of the analysed methods for month-to-month changes and for annual differences is in open conflict with the opinion of many experts. They often believe that month-to-month changes are unpredictable because of erratic cyclical and seasonal fluctuations, while the evolvement of the trend is thought to be stable and, therefore, easier to predict. For the industrial production of Austria and Germany just the opposite seems to be true however. Since correct predictions of the trend play such a prominent role for the forecasting performance, we turn to a closer inspection of this problem before concluding this paper.

Trend change accuracy

Making use of the definitions given in the relations (4) of the theoretical section, we can classify forecasts into downturns / no downturns / upturns / no upturns. Comparing these values with the actual trend movements, we can calculate the percentage of trend changes forecast correctly. These percentages for different forecasting horizons, different forecasting methods and different production countries are given in Table 11. The method, by which these figures are derived is demonstrated exemplarily in Table 12 for one month ahead forecasts with *BSM*.

Table 11

Forecasting performance: trend change error

Percentage of trend changes forecast correctly

Forecasting horizon (months)	Forecasting method	Production country: Austria	Production country: Germany
1	*BSM*	94(4)	93(4)
	SMAC	95(3)	97(1)
	SMACX	97(1)	96(2)
	ARIMA	96(2)	95(3)
6	*BSM*	81(3)	71(3)
	SMAC	83(2)	77(2)
	SMACX	86(1)	79(1)
	ARIMA	80(4)	74(3)
12	*BSM*	64(3)	59(3)
	SMAC	64(3)	62(2)
	SMACX	83(1)	72(1)
	ARIMA	68(2)	58(2)
18	*BSM*	53(3)	44(3)
	SMAC	54(2)	48(2)
	SMACX	79(1)	66(1)
	ARIMA	52(4)	43(4)
24	*BSM*	36(4)	36(3)
	SMAC	42(3)	44(2)
	SMACX	84(1)	63(1)
	ARIMA	44(2)	33(4)

From Table 11, we note immediately that the percentage of trend changes forecast correctly drops significantly with the increasing length of the forecasting horizon. This result is no surprise and corroborates our above hypothesis that trend change errors are a more serious problem than inaccuracies in the prediction of short term components. We detect some differences in the trend change accuracy between Austria and Germany. Apart from the predictions with a forecasting horizon of 1 month, the trend change accuracy for all tested forecasting methods is in Austria substantially better. The most interesting information of this table is the excellent performance of *SMACX* in this context. Splitting up the long run component of a time series into a slowly changing trend component and a repetitive cyclical component leads to a substantial improvement in trend change accuracy. For

Austria, the percentage of trend changes forecast correctly by *SMACX* is, even for a forecasting horizon of 24 months, with more than 80% extremely high. Thus, we observe here a substantial trade-off between absolute and relative forecasting performance, and as a source of qualitative information *SMACX* should not be ignored. This result has a very interesting implication. Apparently, statistical techniques can provide valuable information, if only the right questions are asked. To do this, may not be easy for analysts who are no trained statisticians.

Table 12

Trend change accuracy: one month ahead forecasts with *BSM*

		Austria			Germany	
	Correct	Incorrect	Total	Correct	Incorrect	Total
1 month ahead						
Downturn	2	0	2	2	0	2
No downturn	93	2	95	86	6	92
Total	95	2	97	88	6	94
Percent	97.9	2.1	100.0	93.6	6.4	100.0
Upturn	0	1	1	1	0	1
No upturn	15	4	19	20	2	22
Total	15	5	20	21	2	23
Percent	75.0	25.0	100.0	91.3	8.7	100.0
Overall total	110	7	117	109	8	117
Overall percent	94.0	6.0	100.0	93.2	6.8	100.0

Conclusions

In the present paper, we compare the forecasting performance of structural time series models with that of a traditional *ARIMA* model. The forecasting performance of model can be evaluated from two perspectives. One can use the absolute magnitude of the committed forecast errors as yardstick or one can rely on a model's ability to predict turning points as evaluation criterion. The computed statistics are often refered to as measures of absolute and relative (or directional) accuracy, respectively.

As far as absolute accuracy is concerned, even 24-months-ahead forecasts might be classified as highly accurate according to international standards. Our own standards are more stringent, and we would say that forecasts with a horizon of up to 12 months can be considered as sufficiently reliable. Next, we investigate whether the length of the forecasting horizon, the forecasting method, and the production country have significant effects on the absolute accuracy. The forecasting horizon turns out to be the most influential factor. The longer the horizon, the larger the committed errors. This result is more or less trivial. The forecasting method also has a significant effect. A ranking of the tested methods can be determined. Structural models outperform the traditional *ARIMA* model clearly. Somewhat surprising is however, that the basic structural model does better than more sophisticated model versions with an additive cyclical component. The third tested factor, namely the production country, is of no relevance.

Basing the judgement on relative measures of accuracy changes the ranking of the tested forecasting methods completely. The structural models remain superior, but now the sophicated model versions with additive cyclical component dominate. Here, we also come upon one of the most surprising results of the whole paper. We detect that problems with the prediction of the trend are mainly responsible for a poor directional forecasting performance. Once the trend is removed, the remaining rest can be forecast perfectly, even for a horizon of 24 months.

It follows as final conclusion from this study, that structural time series models are superior to a traditional *ARIMA* model for modelling industrial production in Austria and Germany. Besides a forecasting performance, which is definitely at least as good as that of a *ARIMA* model, they offer additionally valuable information about trend, seasonal and, possibly, cyclical components.

References

Hahn,F.R., Thury,G., Structural Time Series Models for the Austrian and German Industrial Production, Working Paper, Austrian Institute of Economic Research, 49, 1992.

Lewis,C.D., Industrial and Business Forecasting Methods, Butterwoths, London, 1982.

Thury,G., Kalendereffekte in der Industrieproduktion, WIFO-Monatsberichte, 1989, 62(8), pp. 521 - 526.

Witt,S.F., Witt,C.A., Modeling and Forecasting Demand in Tourism, Academic Press, New York, 1992.