Ifo Institute for Economic Research at the University of Munich

ifo Working Papers

Freedom of Choice in Macroeconomic Forecasting: An Illustration with German Industrial Production and Linear Models

Nikolay Robinzonov

Klaus Wohlrabe

Ifo Working Paper No. 57

March 2008

Freedom of Choice in Macroeconomic Forecasting: An Illustration with German Industrial Production and Linear Models*

Abstract

Different studies provide surprisingly a large variety of controversial conclusions about the forecasting power of an indicator, even when it is supposed to forecast the same time series. In this study we aim to provide a thorough overview of linear forecasting techniques and draw conclusions useful for the identification of the predictive relationship between leading indicators and time series. In a case study for Germany we forecast four possible representations of industrial production. Further on we consider a large variety of time-varying specifications: ex post vs. ex ante, rolling vs. recursive and model specifications such as restricted vs. unrestricted, AIC vs. BIC vs. OSC, direct vs. indirect. In a horse race with nine leading indicators plus benchmark we demonstrate the variance of assessment across target variables and forecasting settings (50 per horizon). We show that it is nearly always possible to find situations in which one indicator proved to have better predicting power compared to another.

JEL Code: C52, C53, E37.
Keywords: Forecasting competition, leading indicators, model selection.

Nikolay Robinzonov
Ifo Institute for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich
robinzonov@ifo.de

Klaus Wohlrabe**
Ifo Institute for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich
Phone: +49(0)89/9224-1229
wohlrabe@ifo.de

* We thank Christian Müller, Johannes Mayr, Dirk Ulbricht, and seminar participants in Munich for helpful comments. We are grateful to Jan Grossarth for careful research assistance. The usual disclaimer applies.
** Corresponding Author.

1 Introduction

Why does the forecast performance of one indicator or econometric model prove to be functional in one situation and not in another? It is hard to answer this question, even when the target time series is supposed to be the same. Hüfner and Schröder (2002) found the ZEW Economic Sentiment indicator to have better forecasting properties for German industrial production (yearly growth rates) than its competitor, the Ifo Business Climate. In a replication, Benner and Meier (2004) used monthly growth rates and found opposite results (using a slightly different methodology). On average, in this study the Ifo indicator provided more accurate forecasts than the ZEW indicator.[1] A practitioner asks: How do I forecast a specific macroeconomic time series? The success of macroeconomic forecasts depends either on the choice of a specific econometric model, a specific leading indicator or a combination of both. The out-of-sample forecast is often viewed as the acid test of an econometric model or a leading indicator. "Good" can be assessed in comparison with rival (often naive or other indicators) forecasts. As a practitioner, if you were to look at the empirical literature, you would find a sheer volume of predictor variables under consideration and an endless array of forecasting models and time-varying specifications. Horse races between competing forecasting models and indicators are abundant in the empirical literature. In many cases one can easily encounter a strong correlation between the results and the forecaster's intention. As Denton (1985) notes, often only significant results are ultimately published. A forecaster is confronted with so many different options within the forecasting process. Among these decisions probably the most important point is the employed time series model and its specification. Elliot and Timmermann (2008) review almost all issues concerning economic forecasts. In an empirical application the authors investigate the performance of several time series models models by forecasting inflation and stock returns. Clements and Hendry (1998) illustrate eight dichotomies that intrude on any forecast evaluation exercise. These eight dichotomies relate to the type of model, method of forecasting and forecast evaluation, the nature of economic environment, and the objective of the exercise. In an out-of-sample forecasting exercise they illustrate these dichotomies with a focus on model selection.

Historically, the focus in forecasting has been on low-dimensional univariate or multivariate models all sharing the common linearity in the parameters. In fact, many of the present non-linear techniques are direct generalizations of the linear methods. Recently, there are additional papers that investi-

[1] More details can be found in the literature section.

gate the forecasting performance of non-linear time series models[2] and large scale factor models.[3] Besides the model comparison a focus has been put on assessment of the forecast performance of a leading indicator. Based on the assumption that an indicator and reference (macroeconomic) time series should relate significantly and remain stable, many studies heuristically include some indicators and judge their performance against some others. The variety of verisimilar model estimations is crucial for this judging. In many papers the authors pick out a model and deliver wonderful forecastability results for an indicator and a reference series while suppressing possible other model specifications.

Consequently we ask: Does the forecast performance of a leading indicator depend on the forecasting setting. We conduct a comprehensive study by covering almost all commonly used linear forecasting techniques. This is made by forecasting German industrial production (IP) with nine leading indicators. We demonstrate how the assessment of the forecasting properties may differ between different forecasting settings. Our results are consistent with previous papers on forecasting German industrial production which differ in the assessment of the used indicators.

We use the seasonally adjusted monthly industrial production for Germany. As we want to focus on stationary time series models, we forecast four different stationary representations of the target variable. We calculate four different growth rates, the exact and approximate monthly and yearly growth rates. We consider two different time series models which can be considered as workhorses in forecasting: autoregressive models with exogenous variables (ARX) and vector autoregressive models (VAR). Within these model classes we allow for many different specifications. We distinguish between different model selection criteria. We test whether it makes a difference to employ the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) or an out-of-sample criterion (OSC). Furthermore, we investigate whether a recursive or rolling forecasting scheme is relevant for the assessment of an indicator. We call these many possible forecasting settings freedom of choice in macroeconomic forecasting.

The paper is structured as follows. In section 2 we illustrate the freedom of choice in macroeconomic forecasting. Then we relate this to the existing

[2]See Clements, Franses, and Swanson (2004) for a literature overview, Teräsvirta, van Dijk, and Medeiros (2005) for a recent application of Smooth Transition Autoregressive (STAR) and neural network models, and Claveria, Pons, and Ramos (2007) for an applications of Markov-switching and Self-Exciting Autoregressive (SETAR) models. See Stock and Watson (2003) for a comparison of linear and non-linear time series models.

[3]See Stock and Watson (2002), Forni, Hallin, Lippi, and Reichlin (2003), Dreger and Schumacher (2005), Schumacher (2006), and Eickmeier and Ziegler (2008) among others.

literature for Germany and show how the assessment a of leading indicator can differ across forecasting settings. The empirical results of the comprehensive forecasting competition are presented in section 4. Then we discuss our results and conclude.

2 Methodology - The freedom of choice

In the introduction we presented some options a forecaster is confronted with. In this section we systemize many of them. Our outline is similar to the eight dichotomies presented by Clements and Hendry (1998). We focus on those we want to investigate in our empirical application. At the end of this section we present some more choices an investigator is confronted with. Table 1 displays the different options. We start with describing the data and finish with the different time series models. The freedom of choice is illustrated by 50 possible forecasting settings for a specific time series.

Table 1: Data and Model Considerations

Data	Principle	Method	Model	Restrictions	Selection Criterion
monthly exact	rolling	direct	ARX(p, r)	yes	AIC
monthly approximate	recursive	indirect	VAR(p)	no	BIC
yearly exact					OSC
yearly approximate					

2.1 The data

2.1.1 German Industrial Production

The target variable in our case study is the industrial production (IP) for Germany from 1991:01 to 2006:12. We do not use data before 1991 to circumvent any structural breaks in the data due to the reunification. In order to ensure the same sample size for different specifications we start in 1992:01. The series is seasonally and workday adjusted and was obtained from the Deutsche Bundesbank.[4]
In our case study we only use stationary time series models. Therefore we forecast four stationary representations (interpretations) of the (trending)

[4]Series USNA01.

German industrial production. First we calculate exact monthly and yearly growth rates:

$$\Delta_1 IP = (IP_t - IP_{t-1})/(IP_{t-1}) \tag{1}$$

and

$$\Delta_{12} IP = (IP_t - IP_{t-12})/(IP_{t-12}) \tag{2}$$

and approximate monthly and yearly growth rates calculated as log differences:

$$\tilde{\Delta}_1 IP = \log(IP_t) - \log(IP_{t-1}) \tag{3}$$

and

$$\tilde{\Delta}_{12} IP = \log(IP_t) - \log(IP_{t-12}) \tag{4}$$

For all series we do not remove or change any outliers. In Figure 1 we graph all series. The upper left shows the exact and approximate yearly growth rates, which exhibits a clear periodical pattern. Is is clear that there is only a small difference between exact and approximate growth rates. The differences are graphed in the right panel. Still, it could be interesting if these differences lead to different conclusions. The monthly growth rates display an erratic pattern and seem to be harder to forecast. There are further possible transformations of the target variable, see Marcellino (2006) for examples and references.

The illustration of different interpretations (representations) of a target variable is essential. First, in the literature it is not uncommon to forecast "the GDP" or "the IP" of a specific country but in practice a specific growth rate.[5] Second, the choice of a specific transformation is rarely justified in the literature. In our literature review no article motivates the employed data transformation. And third, as we will show in our case study, the performance and assessment of a leading indicator can differ across different data transformations of the target variable.

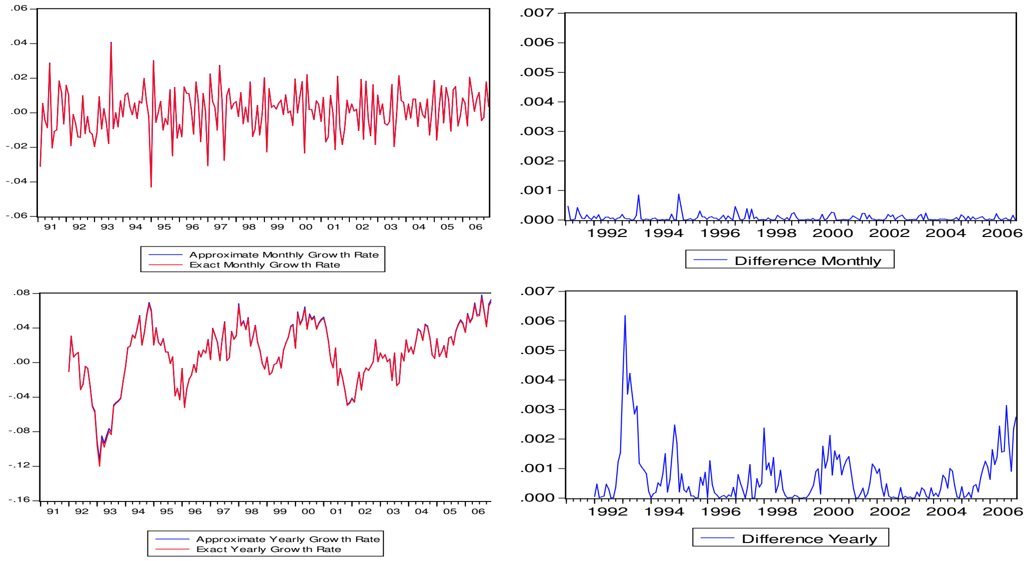
2.1.2 Leading Indicators

In order to illustrate the diversity of forecasting outcomes, we conduct our forecasting exercise with nine leading indicators displayed in Table 2. The choice is guided by the literature on forecasting German IP. For the purpose of illustration it could be any other possible leading indicator combination.

The Ifo Business Climate Index is based on about 7,000 monthly survey responses of firms in manufacturing, construction, wholesaling and retailing.

[5] One could transform the forecast back into the original level series and judge this forecast accuracy but this is not usually done.

Figure 1: Representations of Industrial Production in Germany



The firms are asked to give their assessments of the current business situation and their expectations for the next six months. The balance value of the current business situation is the difference of the percentages of the responses "good" and "poor", the balance value of the expectations is the difference of the percentages of the responses "more favorable" and "more unfavorable". The replies are weighted according to the importance of the industry and aggregated. The business climate is a transformed mean of the balances of the business situation and the expectations. For further information see Goldrian (2007). The ZEW Indicator of Economic Sentiment is surveyed monthly. Up to 350 financial experts take part in the poll. The indicator reflects the difference between the share of analysts that are optimistic and the share of analysts that are pessimistic for the expected economic development in Germany in six months, see Hüfner and Schröder (2002). Compared to the Ifo Index, the overall economy is represented, and macroeconomic factors are expected to be more dominant. The FAZ indicator (Frankfurter Allgemeine Zeitung) pools survey data and macroeconomic time series. It consists of the Ifo Index (0.13), new orders in manufacturing industries (0.56), the real effective exchange rate of the euro (0.06), the interest rate spread (0.08), the stock market index DAX (0.01), the number of job vacancies (0.05) and lagged industrial production (0.11). The Ifo Index, orders in manufacturing and the number of job vacancies enter the indicator equation in levels, while the other variables are measured in first differences. The Early Bird indicator compiled by the Commerzbank also pools different time series and stresses the importance of international business cycles for the German economy. Its components are the real effective exchange rate of the euro (0.35), the short-term real interest rate (0.4) defined as the difference between the short-term nominal rate and core inflation, and the purchasing manager index of U.S. manufactures (0.25). The OECD composite leading indicator is calculated in a more complex way. It is compiled using a modified version of the Phase-Average Trend method (PAT) developed by the US National Bureau of Economic Research (NBER). The indicator is compiled by combining de-trended component series in either their seasonally adjusted or raw form. The component series are selected on the basis of various criteria such as economic significance, cyclical behavior, data quality; timeliness and availability. For Germany the following time series are compiled: orders inflow or demand: tendency (manufacturing) (% balance), Ifo Business Climate Indicator (manufacturing) (% balance), Spread of interest rates (% annual rate), Total new orders (manufacturing), Finished goods stocks: level (manufacturing) (% balance) and Export order books: level (manufacturing) (% balance).

In addition to survey and composite indicators we take some financial indicators as a possible predictors. Since the seminal paper by Estrella and Hardouvelis (1991), financial indicators are more in the focus of forecasting. Stock and Watson (2003) review this literature and conduct a large case study for different OECD countries by forecasting GDP, inflation and industrial production. We selected some indicators from their paper which proved to produce better forecasts for German industrial production than the AR benchmark model. First we start with the growth rate of employment in Germany. As financial indicators we take the overnight interbank interest rate (nominal and real) and a interest spread. For definition see Table 1.

Finally we included a factor obtained from a large data set from Germany. The data set contains German quarterly GDP and 111 monthly indicators from 1992 to 2006.[6] Factor models based on large data sets have received increasing attention in the recent forecasting literature. Factor models aim at finding a few representative common factors underlying a large amount of economic activity. For the US, Stock and Watson (2002) provide evidence for the information content of macroeconomic factors derived from hundreds of macroeconomic time series for future industrial production and inflation.

[6]The estimated factor was kindly provided by Christian Schumacher and is based on the paper Marcellino and Schumacher (2007).

Table 2: Leading Indicators

Indicator	Provider	Label
Ifo Business Climate	Ifo Institute	ifo
ZEW Economic Sentiment	ZEW Institute	zew
Early Bird Indicator	Commerzbank	com
OECD Composite	OECD	OECD
leading indicator for Germany		
FAZ Indicator	FAZ Institute	faz
Employment Growth	Bundesbank	emp
Interest Rate: overnight	IMF	rovnght
Interest rate spread	IMF	rspread
= long term Gov. Bonds − rovnght		
Factor		factor
AR Benchmark		AR

2.2 Principle: Rolling vs. Recursive

The rolling approach makes use of fixed windows of data to re-estimate the parameters over the out-of-sample period, whereas the recursive approach makes use of an increasing window to re-estimate the models. The rolling scheme is relatively attractive when one wishes to guard against moment or parameter drift that is difficult to model explicitly. Without any drifts and breaks an enlarged data base could lead to more precise estimation results and hence better forecasts. Thus a recursive scheme would be preferable.
In our case study the initial forecast date is 2002:01 and the final forecast data is 2006:12 minus the forecast horizon. We forecast 1, 3, 6 and 12 months ahead for each approach. For the rolling forecast the data vintage consists of 120 observations (1992:01 - 2001:12) which is fixed.

2.3 Method: Direct vs. indirect

Forecasts can be generated in two different ways: iterated (indirect or "plug-in") and directly. The iterated forecasts entails first estimating an autoregression, then iterating upon that autoregression to obtain the multiperiod forecast. The direct forecast entails regressing a multiperiod-ahead value of the dependent variable on current and past values of the variable. For example, forecasting the industrial production directly twelve months from now might entail the regression of the IP, twelve months hence, against a constant and the current and past values of IP. In case of iterated forecasts one might include the regression of the IP of the current value on a constant and past

values of IP. Choosing between iterated and direct forecasts involves a trade-off between bias and estimation variance. The iterated method produces more efficient parameter estimates than the direct method, but is prone to bias if the one-step-ahead model is misspecified. See Marcellino, Stock, and Watson (2006) for further details and references. The authors show with a large data set of 170 US monthly macroeconomic time series that iterated forecasts typically outperform the direct forecasts, particulary if long lags of the variables are included in the forecasting models and if the forecast horizon is long. Chevillon and Hendry (2005) and Schorfheide (2005) found that direct multistep forecasts tend to be more accurate in small samples but restrict their conclusions to stationary models under the assumption of some forms of empirical model misspecification.
Eventually the decision between direct and indirect forecasts is an empirical one. For the practitioner the direct seems to be preferable as no assumptions about the future path of the exogenous variable are necessary.

2.4 Method: ex ante vs. ex post

An ex ante forecast is a forecast that uses only information that is available at the forecast origin; it does not use actual values of variables from later periods. In case of iterated multiperiod forecasts one has to forecast the leading indicator for the forecast horizon. Therefore the indicators perform worse just because they are poorly predicted.
In an ex post forecasting setting information from the situation being forecast is employed. The actual values of the causal variables are used, not the forecasted values. This seems in practical applications quite implausible but is justified by the fact that many macroeconomic variables are subject to revisions. So the assumption is not too strong for shorter horizons but could be for longer ones (See Claveria, Pons, and Ramos (2007)).

2.5 Forecasting Models

In this section we briefly outline the two standard linear models used in the empirical forecasting literature. The ARX and the VAR are workhorses in applied forecasting. As we want to focus on the assessment of leading indicators we do not consider pure univariate time series models.
We first consider an ARX(p, r) model that explains the behavior of the endogenous variables as a linear combination of its own and the indicators past

values. The one-step-ahead iterated ARX(p, r) model is given by

$$y_{t+1} = \alpha + \sum_{i=1}^{p} \phi_i y_{t+1-i} + \sum_{j=1}^{r} \theta_j x_{t+1-j} + \varepsilon_t \tag{5}$$

where x_t denotes the (exogeneous) indicator series. For the multistep iterated forecasts we consider two settings. In the ex ante setting we forecast the leading indicator with an AR(p) separately. In the ex post setting we assume that the indicator is known for the forecasting period. The corresponding direct forecast regression is

$$y_{t+h} = \beta + \sum_{i=1}^{p} \delta_i y_{t+1-i} + \sum_{j=1}^{r} \gamma_j x_{t+1-j} + \varepsilon_{t+h} \tag{6}$$

We have to note that we do not allow for a contemporaneous influence of the leading indicator on IP. Direct regressions approaches always produce ex ante forecasts as only information available at that specific time is used. For both model classes we allow a minimum of one lag and a maximum of 12 lags.

We extend the single equation models (5) and (6) to the bivariate case. We consider the following VAR(p) model

$$y_t = \alpha + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t \tag{7}$$

where y_t is now a 2×1 vector containing the IP and the indicator variable, the A_i are fixed (2×2) coefficients matrices, α is a fixed 2×1 vector of intercept terms and finally ε_t is a 2-dimensional white noise process. Again, we allow for a maximum number of 12 lags.

In the sense of Clements and Hendry (1998), ARX models are conditional models, whereas unconditional models endogenize all variables as the VAR.

Apart from the presented linear models non-linear models are used more and more in forecasting macroeconomic time series. Markov-Switching models, smooth-transition autoregressive models (STAR), self-exciting autoregressive models (SETAR), and neural networks, among others are employed in the literature. The results are somewhat mixed. So far it seems that no model class dominates the other ones. See Clements, Franses, and Swanson (2004) for a literature overview, Teräsvirta, van Dijk, and Medeiros (2005) for a recent application of STAR and neural network models, and Claveria, Pons, and Ramos (2007) for a applications of Markov-switching and SETAR models. The inclusion of non-linear models is beyond the scope of this paper. From the practitioners point of view non-linear models are harder to implement compared to standard linear models.

2.6 Model Selection Criterion

Once a specific time series model is chosen it needs to be specified. When deciding on the number of lags included one is faced with a trade-off: Choosing a short lag length might restrict potential intertemporal dynamics and thus yields autocorrelated residuals. Choosing a higher order of lags might however lead to overparameterization problems (overfitting). Due to insufficient degrees of freedom, the model parameters are then imprecisely estimated, yielding large standard errors and high estimation uncertainty.[7] The use of information criteria that build on the likelihood function guarantees the specification of a parsimonious time series model, as they not only reward goodness of fit but include a penalty term, that is an increasing function of the number of estimated parameters. This penalty term thus discourages an overfitting of the system. To give an example: We allow for a maximum of 12 lags. Therefore we estimate a model with 25 parameters in the ARX(p, r) and 50 parameters in the bivariate VAR(p) model (including constants). This gives rise to a risk of overfitting the regression.

We employ two of the most popular selection criteria: the Akaike Information criterion (AIC) and the Baysian information criterion (BIC, sometimes referred to as the Schwarz criterion, SC). The AIC tends to select models that are overparameterized, whereas the BIC is consistent in the sense that as the sample size grows it tends to pick the true model if this model is among the choices. Most researchers apply the BIC criterion because it has performed well in Monte Carlo studies, see e.g. Mills and Prasad (1992). Granger and Jeon (2004) find for a large data set that the BIC criterion tends to select models which have an advantage in forecasting accuracy over the AIC criterion.

In contrast, Granger (1993) pointed out that in-sample selection measures (such as the mentioned standard information criteria) frequently fail to provide strong implications for the out-of-sample performance. Thus, as a third selection criterion we choose an out-of-sample criterion (OSC). The preferred model for each indicator is the one with the lowest mean squared forecast error over the respective forecast horizon. Although this setting requires knowledge beyond the information available in a real time forecasting exercise, the results are sharpened. Especially, they are not biased by some misjudgement concerning the lag parameter in the forecasting models. See Swanson and White (1997) for a systematic investigation concerning out-of-sample model selection.

[7] This is a serious problem in forecasting as it has been shown that high estimation uncertainty is likely to influence adversely the out-of-sample forecast performance of econometric models, see e.g. Fair and Shiller (1990).

Inoue and Kilian (2006) investigate all three criteria for choosing forecasting models.[8] They discuss conditions under which a variety of tools of model selection will identify the model with the lowest true out-of-sample mean squared error among a finite set of forecasting models. They find that selection by AIC and ranking them by recursive MSE yields inconsistent results and have a positive probability of choosing a model which does not have the best forecasting performance while the BIC is consistent for nested models. Elliot and Timmermann (2008) state that consistency is not the most important criterion in forecasting.

2.7 Restrictions

Another issue in model building is the aspect of restrictions. Forecasting with time series models with autoregressive parts can be applied with a restricted or unrestricted parameter space. Consider a VAR model with two variables. In the unrestricted case all parameters up to a specific lag length (chosen by a criterion) are used to make the forecasts. Beyond maximum lag selection the model can still be subject to overfitting. In the easiest restricted case the insignificant parameters are set to zero. We proceed in a different way. We choose a specific lag length, identify the "least" significant parameter, set this value to zero and then reestimate the model. We continue in this fashion until all parameters are significant or at least one parameter is left. In Breitung and Jagodinsky (2001) and Benner and Meier (2004) the restricted forecasts proved to be better than the unrestricted ones.

2.8 Further possible considerations

In terms of the computational burden, the aspect of fixed coefficients vs. updating might be important. The outcome of an empirical forecast comparison exercise can depend on whether model coefficients are continuously updated or are held fixed at in-sample values, especially when there are non-constancies. Models that are robust to location shifts will have a relative advantage for fixed coefficients. In this paper we focus on updating as we pretend to be in an imaginary forecasting situation where an forecast is made independently of the past. Stock and Watson (1996) found evidence of model instability for a large set of macroeconomic and financial variables.
Furthermore an investigator has to account for possible breaks in the data generating process. Clements and Hendry (2006) stress instability as a key

[8]The authors consider only nested models.

determinant of forecasting performance. See Elliot and Timmermann (2008) for references on how to account for these issues.

3 Review of the literature for forecasting industrial production in Germany

As we focus on industrial production (IP) for Germany we review this strand of recent literature. Table 2 describes for every paper how the reference series is constructed, the details for the time series model used, the forecasting approach and horizon and how the forecasts are evaluated. One can see that the approaches distinguish over the different aspects. Although all papers use industrial production as the reference series they are not identical. Besides the article by Fritsche and Stephan (2002) who starts in 1978, all series start in the early 1990s. Almost all employ yearly growth rates (approximate or exact) whereas Benner and Meier (2004) forecast exact monthly growth rates. Given the different target time series it is to be expected that the assessment of indicators varies.[9] All papers apply variations of a VAR model and do a recursive forecasting exercise. As the benchmark model they use an AR process. Due to these differences it is not surprising that the assessment of the indicators turned out to be different from approach to approach, especially for the Ifo and ZEW indicator. Breitung and Jagodzinski (2001) evaluated 30 one-step ahead out-of-sample forecasts within a bivariate VAR model. Considering the unrestricted bivariate VAR model in terms of Theils U the indicators hardly proved to be better than the AR(13) benchmark model. Looking at the restricted VAR (zeroing out insignificant parameters) the results are different. The best indicator is the Early Bird followed by the Ifo indicators. ZEW and FAZ are not able to outperform the benchmark. Fritsche and Stephan (2002) evaluate different sub-indicators of the Ifo Business Climate. The business climate for producer of investment goods and for the manufacturing industry improve the forecast performance compared to an AR benchmark model for a 3 and 6 month horizon. They do not consider the ZEW indicator in their paper. Hüfner and Schröder (2002) compare explicitly the Ifo Business Climate and the ZEW Business Confidence Indicator. Applying the Diebold-Mariano test they find that the ZEW indicator provides for a horizon between 3 and 12 months, significantly better forecasts than the benchmark model (RW). This conclusion cannot be drawn for the Ifo Business Expectations (not Climate). Benner and Meier (2004) respond

[9] It is interesting to note that the authors relate their own results to the previous papers. In a strict sense the competitions are not comparable.

to Hüfner and Schröder (2002) by using the same data set, but they forecast not the yearly growth rate but the monthly growth rate and cast their model in the error correction form. They find that the Ifo Business Expectations provide for any forecast horizon always better forecasts than the ZEW indicator. The results hold both for constant as well as for recursive determined model structure. The difference is not statistically significant. Dreger and Schumacher (2005) conduct both an ex ante and an ex post recursive forecasting exercise. The ZEW indicator provides for all cases a better forecast than the AR benchmark model, but statistically significant is only the 12 month ahead case in the ex post approach. The Ifo indicator performs worse than the ZEW indicator in all cases. Furthermore it does not outperform the benchmark model in any case that is statistically significant. Under perfect foresight the FAZ indicator outperforms the benchmark model at any horizon. This displays completely different results compared to Hüfner and Schröder (2002).

This summary demonstrates some aspects of the freedom of choice in economic forecasting. There is no indicator that dominates across specifications and time series models. A comparison of models or indicators is indeed difficult, as the target time series is not identical. The assessment depends on the definition of time series and forecasting settings.

Article	Reference Series	Indicator	Approach	Forecasting method and horizon	Forecast Evaluation
Breitung and Jagodzinski (2001)	Industrial production, seasonal and work-daily adjusted, approximate yearly growth rates (log differences), 1991:01 - 2001: 06	Ifo, ZEW, COM, FAZ	Bivariate VAR, ex ante, restricted and unrestricted	Recursive, one-step ahead forecasts (1999:01 - 2001:06)	Benchmark model: AR , RMSE, Theils U
Fritsche and Stephan (2002)	Industrial production (excluding construction), approximate yearly growth rates (log differences), 1978:01 - 1998:12	Ifo	Bivariate VAR up to 12 lags, restricted to significant t-values	Recursive (constant lag structure), 3 and 6 months ahead (1991:01 - 1998:12)	Benchmark model: AR, RMSE, Theils U
Hüfner and Schröder (2002)	Industrial production, seasonal adjusted from Deutsche Bundesbank, exact yearly growth rates, 1991:12 - 2000:12	Ifo, ZEW	Bivariate VAR with lag structure obtained from univariate regressions for the dependent variable + BIC for the lags of the other variables	Recursive (constant lag structure), forecast horizon: 1, 3, 6, 9, 12 months, (1994:01 - 2000:09)	Benchmark model: RW RMSE, Theils U, Modified Diebold-Mariano test, Encompassing test
Benner and Meier (2004)	Industrial production, seasonal adjusted from Deutsche Bundesbank, exact, monthly growth rates, 1991:12 - 2000:12, dummies for outliers	Ifo, ZEW	Bivariate vector error correction, dummies for outliers, BIC, restricted to significant t-values	Recursive (constant and recursive lag structure), forecast horizon: 1, 3, 6, 9, 12 months (1994:01 - 2000:09)	Benchmark model: AR RMSE, Theils U, Diebold-Mariono test
Dreger and Schumacher (2005)	Industrial production, seasonal unadjusted, approximate yearly growth rates (log differences), 1992:01 - 2004:12	Ifo, ZEW, FAZ, COM	Bivariate VAR	iterated forecasts, ex ante and ex post approach, OSC criterion (flexible lag structure), Forecast horizon: 1, 3, 6, 9, 12 months (1998:01-2004:12)	Benchmark model: AR , Diebold-Mariano test, pooling forecasts

4 Empirical Results

In our case study we forecast all four representations of German IP 1, 3, 6 and 12 months ahead. We employ the ARX and the VAR models for each of the time series. The lag selection is made via the common information criterions AIC, BIC and OSC (with and without restrictions[10]) and the ARX-specific approaches ex-post and ex-ante. For every specification we conduct the direct and indirect forecasting techniques and finally extend these to both time varying schemes: rolling and the recursive forecasting. In combination these settings sum up to 50 forecasting specifications for each horizon.
With 9 indicators and 4 time series to be forecasted we have 36 possible pairs that are considered for the different forecasting settings mentioned above. Additionally to our nine indicators we forecast each time series with an AR process as a benchmark. On average, a leading indicator should beat such a benchmark model.
In order to demonstrate the variety of assessment of indicators we proceed in four steps. First we outline some general results about the forecasting competition. Second we present the best indicator of each forecasting setting at each horizon. Then we rank all indicators and demonstrate the variance of assessment in an ordinal ranking. Finally, we test all indicators within a specific indicator against each other and examine whether the forecasting errors of one indicator are significantly lower compared to another one. We employ the famous Diebold-Mariano-Test (Diebold and Mariano (1995)) with the small sample correction proposes by Harvey, Leybourne, and Newbold (1997).

4.1 General Remarks

As general results we can state that ARX models perform better, on average, than VAR models. This result is interesting because no paper in our literature review for Germany considers ARX models. There could be several explanations for this. First, in the ex post setting we assume the indicator to be known for the forecasting period. This provides information beyond the forecasting date and may result in more accurate forecasts. Second, in a (iterative) VAR setting both variables are forecasted with its own past values and the other variable. This can introduce higher forecasting errors because the leading indicator is supposed to forecast the target variable and vice versa.
Furthermore in about 70% of the cases the rolling scheme produces lower

[10] Due to very large possible lag combinations we abstain from restricted forecasts in the OSC case.

RMSE than the corresponding recursive scheme. The OSC criterion delivers by far lower RMSEs than the statistical selection criterions AIC and BIC. Finally, the indirect approach seems to outperform the direct approach. These results are similar to previous findings in the literature.

4.2 The winners

For each setting and indicator we calculate the Root Mean Squared Error (RMSE). Tables 3 to 6 present the best indicator for each model specification chosen by the lowest RMSE. For the exact yearly growth rates, the AR benchmark model and the *factor* are the dominant winners. But also the financial indicators (*rspread* and *rovnght*), the Early Bird indicator (*com*) and the FAZ indicator (*faz*) provide the lowest RMSE in some specifications. Using the direct approach, the AR benchmark can hardly be outperformed by an indicator. Furthermore, we can state that there is almost no difference between the rolling and the recursive forecasting scheme. In almost all situations we find for both schemes the same indicator with the lowest RMSE. Comparing exact and approximate yearly growth rates (Table 4), we find some slight changes. In 23 cases out of 200 possible settings we find a different winner in a specific forecasting setting. This points to the fact that there are differences in the assessment of indicators across different transformations of the target variable even if the differences are very small (Figure 1).

If we look at the exact monthly growth rates (Table 5), we find a heterogeneous picture. Across forecast horizons and settings you can find a situation where one indicator of our selection provides the lowest RMSE. Considering $h = 6$ all indicators are winners in one or more forecasting settings. Hardly any indicator can outperform the AR benchmark within the direct ARX settings. In 22 cases we find different winners by comparing exact and approximate monthly growth rates (Table 6).

4.3 The ordinal ranking

Tabulating only the winners for each setting does not yield an impression of the variance of assessment over different forecasting settings. Figures 2 to 5 display the corresponding ranking boxplots. For each time series, horizon and indicator we draw the boxplot over all 50 forecasting settings and rankings. For the exact yearly growth rates (Figure 2), we see that *factor* indicator is relatively robust across settings and horizons. The *faz* indicator performs well on shorter horizon and gets comparably worse for longer horizons. We can state the same results for the AR benchmark. For shorter horizon (1 and 3) the benchmark model is difficult to beat whereas for longer horizons the indicators seem to have more information content for forecasting than the pure autoregressive part. On average, the Ifo indicator *ifo* is the worst one especially in the short run. But we have to note that the Ifo Business Climate is constructed for the whole economy and not specifically for the industry sector.[11] In assessing an indicator one has to be aware of the fact that indicators are often constructed to forecast (or describe coincidentally) a specific target variable. It is therefore no surprise that the *faz* performs well because it is constructed to lead industrial production. Comparing the graphs for approximate and exact growth rates, we can find some small differences, i.e. that the ranking of indicators is different for these two target variables. In the case of the monthly growth rates, the ranking variance is much more pronounced (Figure 4 and 5). It is always possible by comparing two specific indicators to find a forecasting setting where one indicator is better than another and vice versa. For illustration purposes we consider the first and hence the most accurate horizon of the exact monthly growth rates and compare two indicators: *faz* and *com*. Under a direct ARX model with recursive time varying scheme and simple AIC criterion *com* outperforms all indicators while *faz* is evaluated as the worst one. For the same horizon and times series under direct VAR model with rolling time varying scheme and BIC criterion, the *com* indicator strongly deteriorates to the ninth place while *faz* becomes the winner. The same large quality magnitude can be found between *faz* and *rrovnght* for the exact yearly growth rates as well. Again, we find small differences in the ordinal ranking between exact and approximate growth rates.

4.4 Statistical Tests

The ordinal ranking does not reveal any significant differences in the forecasting quality. One can obtain a ranking where the RMSEs are so close together that all indicators can be assumed to be suitable in this specific forecasting setting. In order to get a qualitative assessment we test whether the RMSEs of two competing indicators within a specific setting are significantly different. We employ the famous Diebold-Mariano-Test with the small sample correction proposed by Harvey, Leybourne, and Newbold (1997). We test whether the RMSE obtained by one indicator is significantly lower

[11] It contains also survey information from the construction, wholesale and retail sector. We repeated the exercise for the Business Climate for the industry sector (which can be obtained from the Ifo Institute). The results were much better and can be obtained from the authors upon request.

than the RMSE obtained by the competitor. Therefore our null hypothesis: $H_0 = E[\delta_t] \leq 0$, where the sequence of loss differentials δ_t is definded by: $\delta_t = g(e_{it}) - g(e_{jt})$. The loss functions $g(e_{it})$ and $g(e_{jt})$ are derived from the forecast errors e_{it} and e_{jt}. Under the null, model i provides significantly better forecasts than model j. In each forecasting setting we test all indicators (including the AR benchmark) against each other.

In Tables 7 to 10 we display the results. The tables can be read as follows: in each column one can read the number of winners, i.e. an indicator proved to have significantly lower RMSE compared to the competitor. Line by line we count the opposite results, i.e. how often an indicator obtained significantly higher RMSEs. To give an example: Consider in Table 7 the *ifo* and the *zew* indicator. For $h = 1$, in 17 forecasting settings the *zew* provided significantly lower RMSEs compared to *ifo*. Testing the other direction the *ifo* provided in no situation significantly lower RMSEs compared to *zew*. In 33 forecasting settings (out of 50) there were no significant differences between the two indicators. In general the tables confirm the visual impressions form ordinal ranking. In case the exact yearly growth rate AR benchmark and the factor indicator dominate its competitors. A notable exception is the Early Bird indicator (*com*). It never performs worse than the AR benchmark across all settings and horizons. The dominance of the *factor* indicator and AR benchmark deteriorates with increasing forecast horizon. For $h = 12$ we can sharpen our statement from the previous setting. It is not only possible to find a setting where one indicator is ranked better than another one; one can also find a setting where it is significantly better.

Tables 8 and 9 present the results for the monthly growth rates. Again, *factor* and AR dominate the short term forecast assessments. Compared to the yearly growth rates less comparisons between indicators provide significantly different RMSEs. For $h = 12$ for exact yearly growth rates in 795 cases one indicator was significantly better than another, whereas for monthly growth rates we have only 548 cases.

5 Implications for assessing forecasts

Given the empirical results we obtained through our case study, we make some cautionary notes before calculating and assessing forecasts or the performance of leading indicators.

1. Transformation of the target variable.

In macroeconomic forecasting, inflation, Gross Domestic Product (GDP) and IP are the most forecasted variables. Especially the latter two are often transformed into stationary representations for statistical reasons. One common approach is to calculate growth rates or taking first differences. Growth rates can be firstly calculated monthly, quarterly or yearly or secondly exact or approximate (log differences). We have shown in our case study that the performance of forecasting models and indicators differ across different data transformations. Furthermore, the length of the time series should be justified.

2. Care in model selection.
Within a model class, selecting a specific model is a complex task. We have demonstrated easier ways to find a forecasting model (AIC and BIC). We omit complex model selection procedures like the general-to-specific or specific-to-general approach. Given a specific loss function, a forecaster has to decide whether he prefers a statistically "correct" model or a model with the best forecast performance (OSC), which rarely coincide.

3. Considering a large class of possible indicators.
In our case study we considered nine possible indicators for German industrial production. There are many other possible indicators, i.e. financial ones. This can help to identify a possible robust indicator across different forecasting settings. By comparing indicators, the choice can be made on a broader range of indicators. Furthermore, one should be aware of the fact that indicators are constructed differently and contain different information. Thus an indicator is supposed to have better forecast performance for another time series. The poor performance of the Ifo Business Climate can be explained by the fact as it is constructed for the economy as a whole and not just for industrial production as the FAZ indicator. The Ifo Institute also provides a Business Climate indicator for the industry sector which is not commonly known to the public.

4. Given the large information set: Robustness and forecast combinations.
By promoting many model classes and model selection procedures the forecast can be ground on a more deepened and robust scientific basis compared to employing only one specific forecasting setting. Forecasts comparisons are essential under many forecasting settings. Stock and Watson (1999) rank forecasting procedures over a wider range of data sets and see which ones perform well on average. It allows a researcher to identify a possible robust indicator or model specification. A method called 'data snooping' by White (2000) can be employed when a large set of models needs to be compared. This method compares a set of risk estimates generated by a range of indi-

vidual forecasts to the risk of a benchmark model. The null hypothesis is that the best of the forecasting methods is not better than the benchmark models.
Suppose that one model seems to be better than another, it is not clear that it is optimal to ignore the forecasts from the weaker model altogether. Despite the numerous attempts to choose a single forecasting model, empirically it seems that combining forecasts from multiple models with different indicators often outperforms forecasts from a single model, see Makridakis and Hibon (2000) or Marcellino (2004) for empirical applications. A key issue in forecast combination is how the weight is assigned to the various forecasts. See Timmermann (2006) for methods and references.

5. Report of comprehension results.
Given the previous points and our literature review for forecasting German industrial production leads us to point out that horse races between models and indicators should be enlarged and interpreted on a much wider information basis. It would make the results more objective.

6. Comparability of results.
As no paper can include any possible forecasting setting, the comparability of results is important to make progress in forecasting. The comparability is assured when the same interpretation (representation) of a macroeconomic time series is used. Furthermore the time series length should be the same.[12]

6 Conclusions

In this paper we illustrated the freedom of choice in macroeconomic forecasting. By this we mean that a forecaster can decide in favor of so many specifications within the forecasting process that the assessment of forecasting models and leading indicators varies across forecasting settings. We illustrate this freedom of choice in a comprehensive case study by forecasting German industrial production and linear time series models. We employ the two workhorse models mentioned in the literature, the ARX and the VAR model. Within these two model classes we allow for different model selection criteria, the AIC, the BIC and an out-of-sample criteria. Furthermore, we allow for restrictions on insignificant lags. We distinguish between ex post

[12]One first step are the data sets described in Makridakis and Hibon (2000) (http://www.forecastingprinciples.com/m3-competition.html)
and Croushore and Stark (2001) ((http://www.philadelphiafed.org/econ/forecast/real-time-data/index.cfm). They are available on the internet.

and ex ante set ups, direct and indirect forecasts and a rolling and recursive forecasting scheme. Finally we have 50 possible forecasting settings for each horizon. We forecast four representations of industrial production, monthly and yearly growth rates (exact and approximate). In a horse race we compare the forecast performance of nine leading indicators plus AR benchmark for each time series and forecasting setting. Our results show that there is a large variance of the assessment across indicators and forecast settings. It is nearly always possible to find situations where one indicator is (significantly) better than another and vice versa.
Given our results we recommend expanding the information basis for decisions based on forecasts, i.e. considering more model classes, indicators and model selection processes. This would probably allow the forecaster to identify robust models or indicators. Moreover, this facilitates and establishment of rich forecasting combinations, which can be better than single forecasts.

References

Benner, J., and C. Meier (2004): "Prognosegüte alternativer Frühindikatoren für die Konjunktur in Deutschland," *Jahrbücher für Nationalökonomie und Statistik*, 224(6), 637–652.

Breitung, J., and D. Jagodzinski (2001): "Prognoseeigenschaften alternativer Indikatoren der konjunkturellen Entwicklung in Deutschland," *Konjunkturpolitik*, 47(4), 292–314.

Chevillon, G., and D. Hendry (2005): "Non-parametric direct multistep estimation for forecasting economic processes," *International Journal of Forecasting*, 21(2), 201–218.

Claveria, O., E. Pons, and R. Ramos (2007): "Business and consumer expectations and macroeconomic forecasts," *International Journal of Forecasting*, 23(1), 47–69.

Clements, M., P. Franses, and N. Swanson (2004): "Forecasting economic and financial time-series with non-linear models," *International Journal of Forecasting*, 20(2), 169–183.

Clements, M., and D. Hendry (1998): "Evaluating a Model by Forecast Performance," *Oxford Bulletin of Economics and Statistics*, 67, S931–S956.

——— (2006): "Forecasting with Breaks in Data Processes," in *Handbook of Economic Forecasting*, ed. by C. Granger, G. Elliot, and A. Timmermann, pp. 605–657. Amsterdam, North Holland.

Croushore, D., and T. Stark (2001): "A real-time data set for macroeconomists," *Journal of Econometrics*, 105(1), 111–130.

Denton, F. (1985): "Data Mining as an Industry," *The Review of Economics and Statistics*, 67(1), 124–127.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business Economic Studies*, 13(3), 253–263.

Dreger, C., and C. Schumacher (2005): "Out-of-sample Performance of Leading Indicators for the German Business Cycle. Single vs Combined Forecasts," *Journal of Business Cycle Measurement and Analysis*, 2(1), 71–88.

Eickmeier, S., and C. Ziegler (2008): "How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach," *Journal of Forecasting*, forthcoming.

Elliot, G., and A. Timmermann (2008): "Economic Forecasting," *Journal of Economic Literature*, Forthcoming.

Estrella, A., and G. Hardouvelis (1991): "The Term Structure as a Predictor of Real Economic Activity," *The Journal of Finance*, 46(2), 555–576.

Fair, R., and R. Shiller (1990): "Comparing Information in Forecasts from Econometric Models," *The American Economic Review*, 80(3), 375–389.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003): "Do financial variables help forecasting inflation and real activity in the euro area?," *Journal of Monetary Economics*, 50(6), 1243–1255.

Fritsche, U., and S. Stephan (2002): "Leading Indicators of German Business Cycles - An Assessment of Properties," *Jahrbücher für Nationalökonomie und Statistik*, 222(3), 289–311.

Goldrian, G. (2007): *Handbook of survey-based business cycle analysis.* Edward Elgar.

Granger, C. (1993): "On the Limitations of Comparing Mean Squared Forecast Errors: Comment," *Journal of Forecasting*, 12(8), 651–652.

Granger, C., and Y. Jeon (2004): "Forecasting Performance of Information Criteria with Many Macro Series," *Journal of Applied Statistics*, 31(10), 1227–1240.

Harvey, D., S. Leybourne, and P. Newbold (1997): "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 13(2), 281–291.

Hüfner, F., and M. Schröder (2002): "Prognosegehalt von ifo-Geschäftserwartungen und ZEW-Konjunkturerwartungen: Ein ökonometrischer Vergleich," *Jahrbücher für Nationalökonomie und Statistik*, 222(3), 316–336.

Inoue, A., and L. Kilian (2006): "On the Selection of Forecasting Models," *Journal of Econometrics*, 130(2), 273–306.

Makridakis, S., and M. Hibon (2000): "The M3-Competition: results, conclusions and implications," *International Journal of Forecasting*, 16(4), 451–476.

Marcellino, M. (2004): "Forecast Pooling for European Macroeconomic Variables," *Oxford Bulletin of Economics and Statistics*, 66(1), 91–112.

——— (2006): "Leading Indicators," in *Handbook of Economic Forecasting*, ed. by C. Granger, G. Elliot, and A. Timmermann, pp. 879–960. Amsterdam, North Holland.

Marcellino, M., and C. Schumacher (2007): "Factor nowcasting of German GDP with ragged-edge data. A model comparison using MIDAS projections," Discussion paper, Bundesbank Discussion Paper, Series 1, 34/2007.

Marcellino, M., J. Stock, and M. Watson (2006): "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics*, 135(1-2), 499–526.

Mills, J., and K. Prasad (1992): "A comparison of model selection criteria," *Econometric Reviews*, 11(2), 201–234.

Schorfheide, F. (2005): "VAR forecasting under misspecification," *Journal of Econometrics*, 128(1), 99–136.

Schumacher, C. (2006): "Forecasting German GDP using alternative factor models based on large datasets," *Journal of Forecasting*, 26(4).

STOCK, J., AND M. WATSON (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14(1), 11–30.

——— (1999): "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, ed. by R. Engle, and H. White, pp. 879–960. Oxford University Press.

——— (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20(2), 147–62.

——— (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41(3), 788–829.

SWANSON, N., AND H. WHITE (1997): "Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models," *International Journal of Forecasting*, 13(4), 439–61.

TERÄSVIRTA, T., D. VAN DIJK, AND M. MEDEIROS (2005): "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A reexamination," *International Journal of Forecasting*, 21(4), 755–774.

TIMMERMANN, A. (2006): "Forecast Combinations," in *Handbook of Economic Forecasting*, ed. by C. Granger, G. Elliot, and A. Timmermann, pp. 135–196. Amsterdam, North Holland.

WHITE, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68(5), 1097–1126.

Table 3: Forecast Competition - Exact Yearly Growth Rates - Best Indicator

	Horizon →	1		3		6		12	
Model	Criterion ↓	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling
ARX(p,r)	AIC	factor	factor	factor	factor	factor	factor	com	factor
unrestricted	BIC	AR	faz	factor	factor	factor	factor	factor	factor
indirect, ex ante	OSC	rovnght	rovnght	rspread	rspread	rspread	rspread	rspread	rspread
ARX(p,r)	AIC	AR	factor	factor	factor	factor	factor	com	factor
restricted	BIC	AR	faz	factor	factor	factor	factor	factor	factor
indirect, ex ante									
ARX(p,r)	AIC	factor	factor	factor	factor	faz	factor	faz	factor
unrestricted	BIC	AR	faz	factor	faz	faz	faz	faz	faz
indirect, ex post	OSC	rspread	rspread	rspread	rspread	rspread	rspread	rspread	rspread
ARX(p,r)	AIC	AR	factor	faz	factor	faz	factor	faz	faz
restricted	BIC	AR	faz	factor	faz	faz	faz	faz	faz
indirect, ex post									
ARX(p,r)	AIC	AR	AR	AR	AR	AR	AR	AR	AR
unrestricted	BIC	AR	AR	AR	AR	AR	AR	com	AR
direct, ex ante	OSC	AR	emp	emp	emp	emp	emp	rspread	emp
ARX(p,r)	AIC	AR	AR	AR	AR	AR	AR	com	AR
restricted	BIC	AR	faz	AR	AR	AR	AR	com	AR
direct, ex ante									
Bivariate VAR(p)	AIC	factor	factor	factor	factor	factor	factor	rovnght	rovnght
unrestricted	BIC	AR	AR	AR	AR	factor	factor	rovnght	factor
indirect, ex ante	OSC	faz	factor	factor	faz	faz	factor	faz	rovnght
Bivariate VAR(p)	AIC	AR	factor	factor	factor	factor	factor	rovnght	rovnght
restricted	BIC	AR	AR	AR	factor	rovnght	factor	rovnght	factor
indirect, ex ante									
Bivariate VAR(p)	AIC	factor	factor	AR	AR	AR	AR	AR	AR
unrestricted	BIC	AR	AR	AR	AR	rovnght	AR	com	com
direct, ex ante	OSC	faz	factor	faz	faz	faz	factor	factor	factor
Bivariate VAR(p)	AIC	AR	factor	AR	AR	AR	AR	AR	AR
restricted	BIC	AR	AR	AR	AR	rovnght	AR	com	com
direct, ex ante									

Table 4: Forecast Competition - Approximate Yearly Growth Rates - Best Indicator

	Horizon →	1		3		6		12	
Model	Criterion ↓	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling
ARX(p, r)	AIC	factor	factor	factor	factor	factor	factor	com	factor
unrestricted	BIC	AR	faz	factor	factor	factor	factor	factor	factor
indirect, ex ante	OSC	rovnght	rovnght	rspread	rspread	rspread	rspread	rspread	rspread
ARX(p, r)	AIC	factor	factor	factor	factor	factor	factor	com	factor
restricted	BIC	factor	AR	factor	factor	factor	factor	factor	factor
indirect, ex ante									
ARX(p, r)	AIC	factor	factor	factor	factor	faz	factor	faz	factor
unrestricted	BIC	AR	faz	factor	faz	faz	faz	faz	faz
indirect, ex post	OSC	com	rovnght	rspread	rspread	rspread	rspread	rspread	rspread
ARX(p, r)	AIC	factor	factor	factor	factor	faz	faz	faz	faz
restricted	BIC	factor	AR	factor	faz	faz	faz	faz	faz
indirect, ex post									
ARX(p, r)	AIC	AR	AR	AR	AR	AR	AR	AR	AR
unrestricted	BIC	AR	AR	AR	AR	AR	AR	com	AR
direct, ex ante	OSC	emp	emp	emp	emp	emp	emp	emp	emp
ARX(p, r)	AIC	AR	AR	AR	AR	AR	AR	com	AR
restricted	BIC	AR	faz	AR	AR	AR	AR	com	AR
direct, ex ante									
Bivariate VAR(p)	AIC	factor	factor	factor	factor	factor	factor	rovnght	rovnght
unrestricted	BIC	AR	AR	AR	AR	rovnght	factor	rovnght	factor
indirect, ex ante	OSC	faz	faz	factor	faz	faz	faz	faz	rovnght
Bivariate VAR(p)	AIC	factor	factor	factor	factor	factor	factor	rovnght	rovnght
restricted	BIC	AR	AR	AR	AR	rovnght	factor	rovnght	com
indirect, ex ante									
Bivariate VAR(p)	AIC	factor	factor	AR	AR	AR	AR	AR	AR
unrestricted	BIC	AR	AR	AR	AR	rovnght	AR	com	com
direct, ex ante	OSC	faz	faz	faz	faz	factor	factor	factor	factor
Bivariate VAR(p)	AIC	factor	factor	AR	AR	AR	AR	com	AR
restricted	BIC	AR	AR	AR	AR	rovnght	AR	com	com
direct, ex ante									

Table 5: Forecast Competition - Exact Monthly Growth Rates - Best Indicator

	Horizon →	1		3		6		12	
Model	Criterion ↓	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling
ARX(p, r)	AIC	factor	factor	rspread	rspread	oecd	rovnght	rovnght	rovnght
unrestricted	BIC	factor	factor	rovnght	zew	com	oecd	rovnght	rovnght
indirect, ex ante	OSC	AR	AR	AR	AR	rovnght	rovnght	emp	emp
ARX(p, r)	AIC	factor	factor	rspread	rspread	oecd	rspread	rovnght	rspread
restricted	BIC	factor	factor	rovnght	zew	com	oecd	com	zew
indirect, ex ante									
ARX(p, r)	AIC	factor	factor	rspread	rspread	emp	rspread	emp	rspread
unrestricted	BIC	factor	factor	emp	emp	faz	emp	faz	emp
indirect, ex post	OSC	AR	AR	AR	AR	rspread	rspread	ifo	ifo
ARX(p, r)	AIC	factor	factor	rspread	rspread	rspread	rspread	rspread	rspread
restricted	BIC	factor	factor	rovnght	rspread	faz	com	faz	com
indirect, ex post									
ARX(p, r)	AIC	AR	AR	AR	AR	AR	AR	AR	AR
unrestricted	BIC	AR	AR	AR	AR	AR	AR	rovnght	AR
direct, ex ante	OSC	AR	AR	AR	AR	emp	emp	emp	emp
ARX(p, r)	AIC	AR	AR	AR	AR	AR	AR	AR	AR
restricted	BIC	AR	AR	AR	AR	AR	AR	rovnght	AR
direct, ex ante									
Bivariate VAR(p)	AIC	factor	factor	rspread	rspread	rspread	rovnght	rovnght	rovnght
unrestricted	BIC	faz	faz	faz	factor	faz	ifo	rovnght	rovnght
indirect, ex ante	OSC	faz	faz	faz	faz	faz	faz	faz	faz
Bivariate VAR(p)	AIC	rspread	rspread	rspread	rspread	rspread	rspread	rspread	zew
restricted	BIC	factor	rspread	rovnght	ifo	zew	zew	rovnght	zew
indirect, ex ante									
Bivariate VAR(p)	AIC	factor	factor	rovnght	AR	factor	AR	rspread	AR
unrestricted	BIC	faz	faz	rovnght	rovnght	rovnght	AR	rovnght	AR
direct, ex ante	OSC	faz	faz	faz	faz	faz	faz	faz	oecd
Bivariate VAR(p)	AIC	rspread	rspread	zew	zew	zew	AR	rspread	AR
restricted	BIC	factor	rspread	zew	zew	rovnght	AR	rovnght	AR
direct, ex ante									

Table 6: Forecast Competition - Approximate Monthly Growth Rates - Best Indicator

	Horizon →	1		3		6		12	
Model	Criterion ↓	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling	Recursive	Rolling
ARX(p, r)	AIC	factor	factor	rspread	rspread	oecd	rspread	rovnght	rovnght
unrestricted	BIC	factor	factor	faz	zew	com	oecd	rovnght	rovnght
indirect, ex ante	OSC	AR	AR	AR	AR	rovnght	rovnght	emp	emp
ARX(p, r)	AIC	rspread	rspread	rspread	rspread	oecd	rspread	rovnght	rspread
restricted	BIC	zew	zew	rovnght	zew	com	zew	com	zew
indirect, ex ante									
ARX(p, r)	AIC	factor	factor	rspread	rspread	emp	rspread	emp	rspread
unrestricted	BIC	factor	factor	rovnght	emp	faz	emp	faz	emp
indirect, ex post	OSC	AR	AR	AR	AR	rspread	rspread	ifo	ifo
ARX(p, r)	AIC	rspread	rspread	rspread	rspread	rspread	rspread	rspread	rspread
restricted	BIC	zew	zew	rovnght	com	faz	com	faz	com
indirect, ex post									
ARX(p, r)	AIC	AR	AR	AR	AR	AR	AR	AR	AR
unrestricted	BIC	AR	AR	AR	AR	AR	AR	rovnght	AR
direct, ex ante	OSC	AR	AR	AR	AR	emp	emp	emp	emp
ARX(p, r)	AIC	AR	AR	AR	AR	AR	AR	AR	AR
restricted	BIC	AR	AR	AR	AR	AR	AR	rovnght	AR
direct, ex ante									
Bivariate VAR(p)	AIC	factor	factor	rspread	rspread	rspread	rovnght	rovnght	rovnght
unrestricted	BIC	faz	faz	rovnght	factor	faz	ifo	rovnght	rovnght
indirect, ex ante	OSC	faz	faz	faz	faz	faz	faz	faz	faz
Bivariate VAR(p)	AIC	rspread	rspread	rspread	rspread	rspread	rspread	rspread	zew
restricted	BIC	faz	zew	rovnght	ifo	rovnght	zew	rovnght	zew
indirect, ex ante									
Bivariate VAR(p)	AIC	factor	factor	rovnght	AR	factor	AR	AR	AR
unrestricted	BIC	faz	faz	rovnght	rovnght	rovnght	AR	rovnght	AR
direct, ex ante	OSC	faz	faz	faz	faz	faz	faz	faz	ifo
Bivariate VAR(p)	AIC	rspread	rspread	zew	zew	factor	AR	rspread	AR
restricted	BIC	faz	zew	zew	zew	rovnght	AR	rovnght	AR
direct, ex ante									

Figure 2: Ranking of leading indicators: Exact yearly growth rates

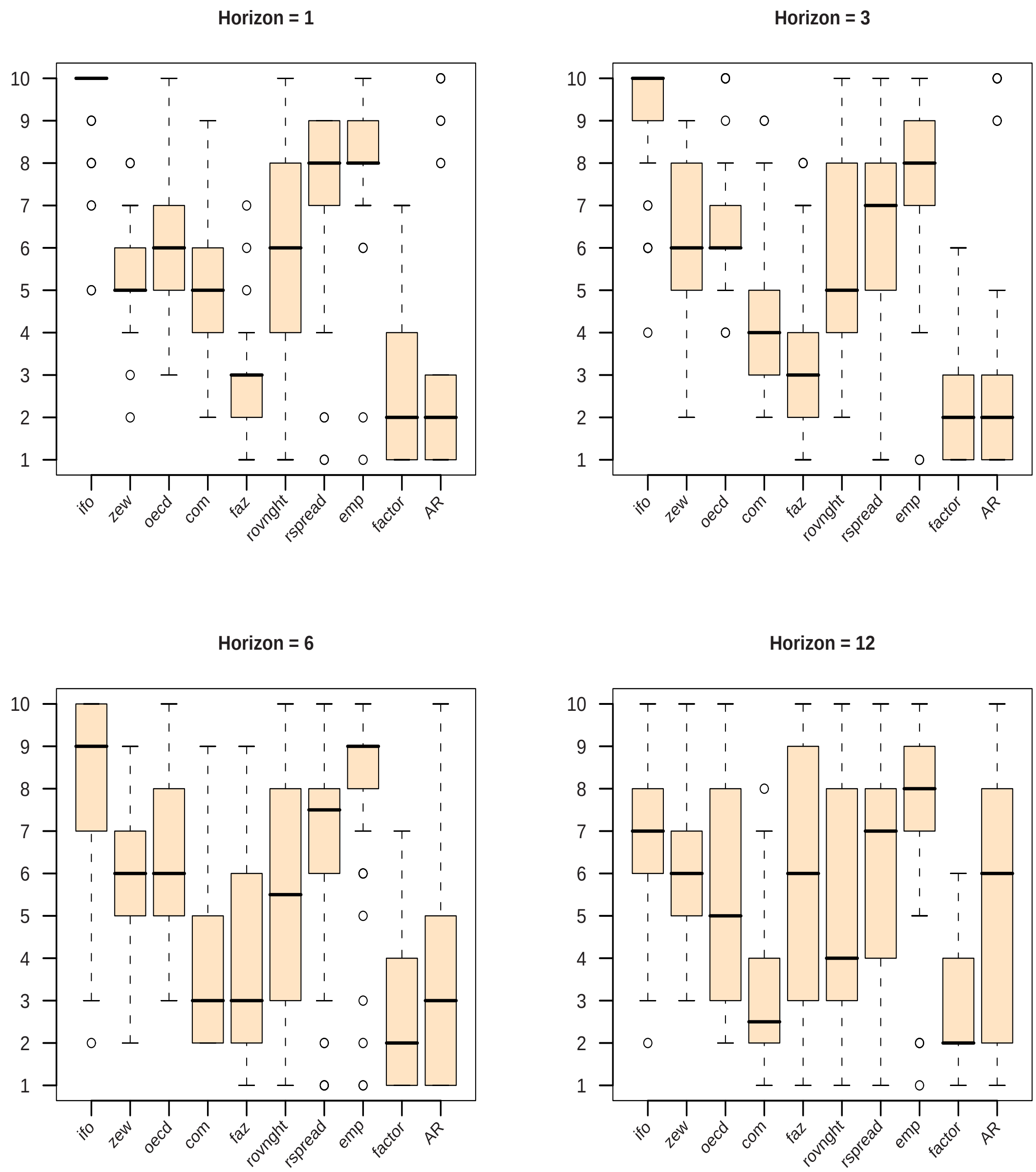


Figure 3: Ranking of leading indicators: Approximate yearly growth rates

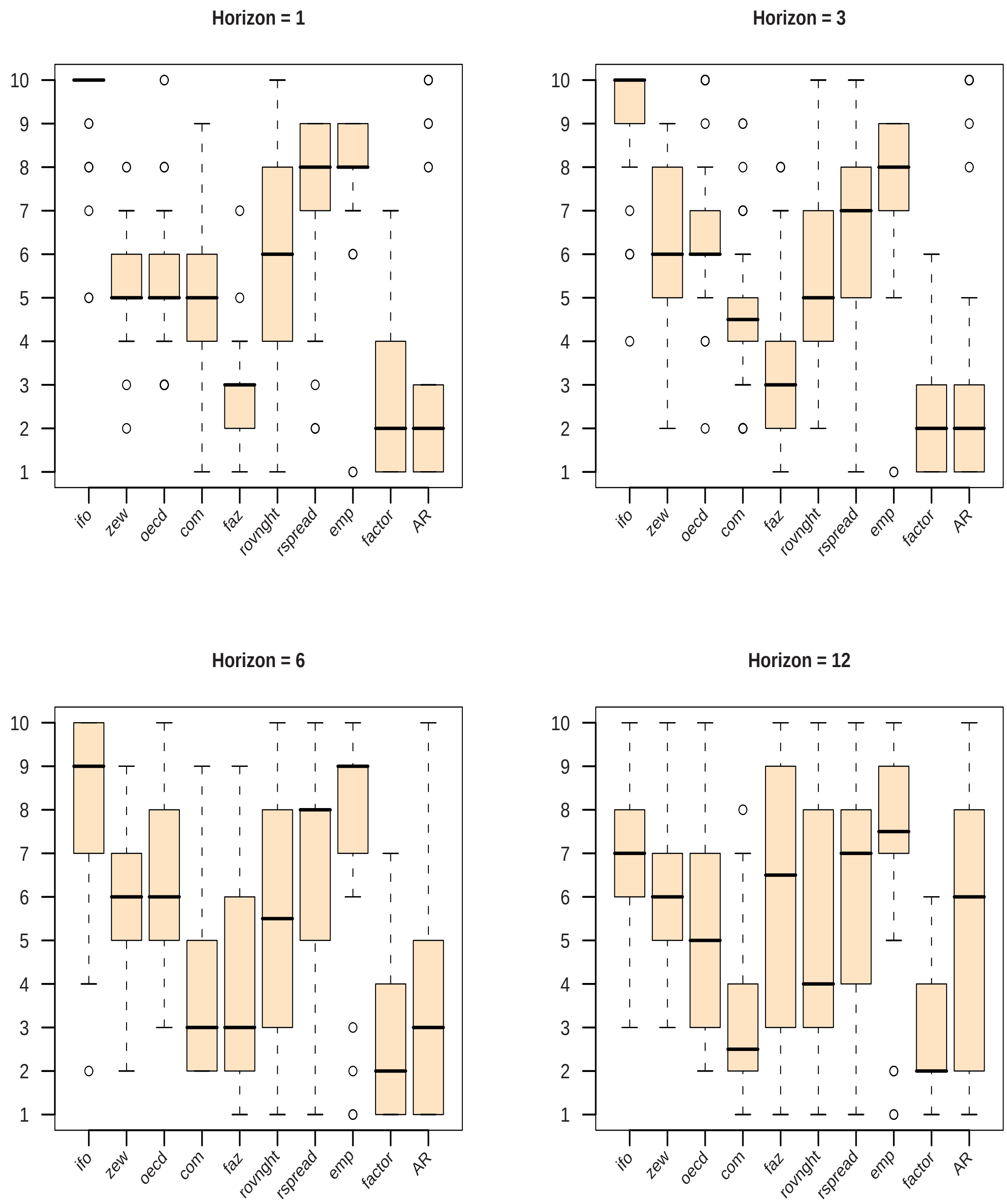


Figure 4: Ranking of leading indicators: Exact monthly growth rates

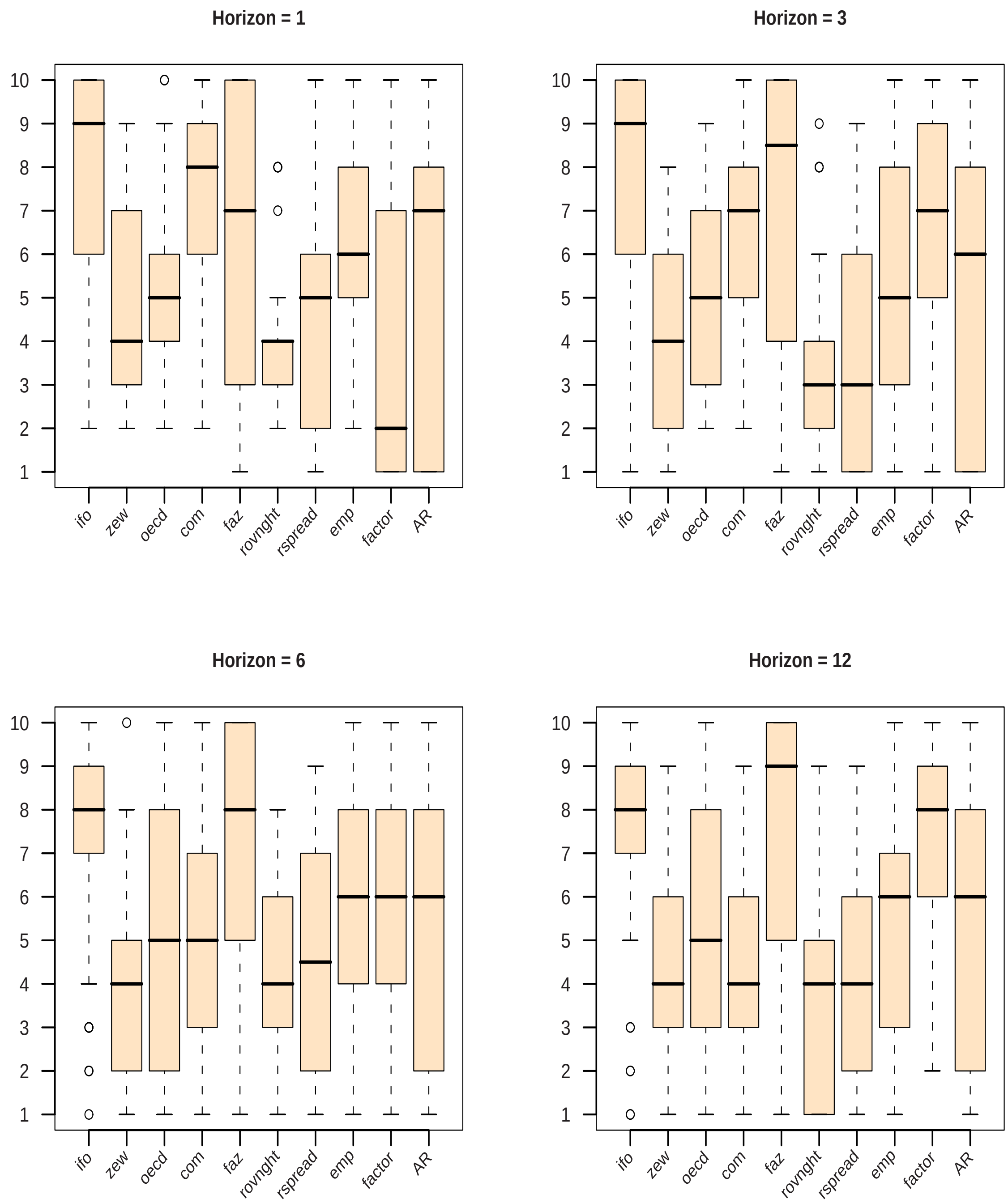


Figure 5: Ranking of leading indicators: Approximate monthly growth rates

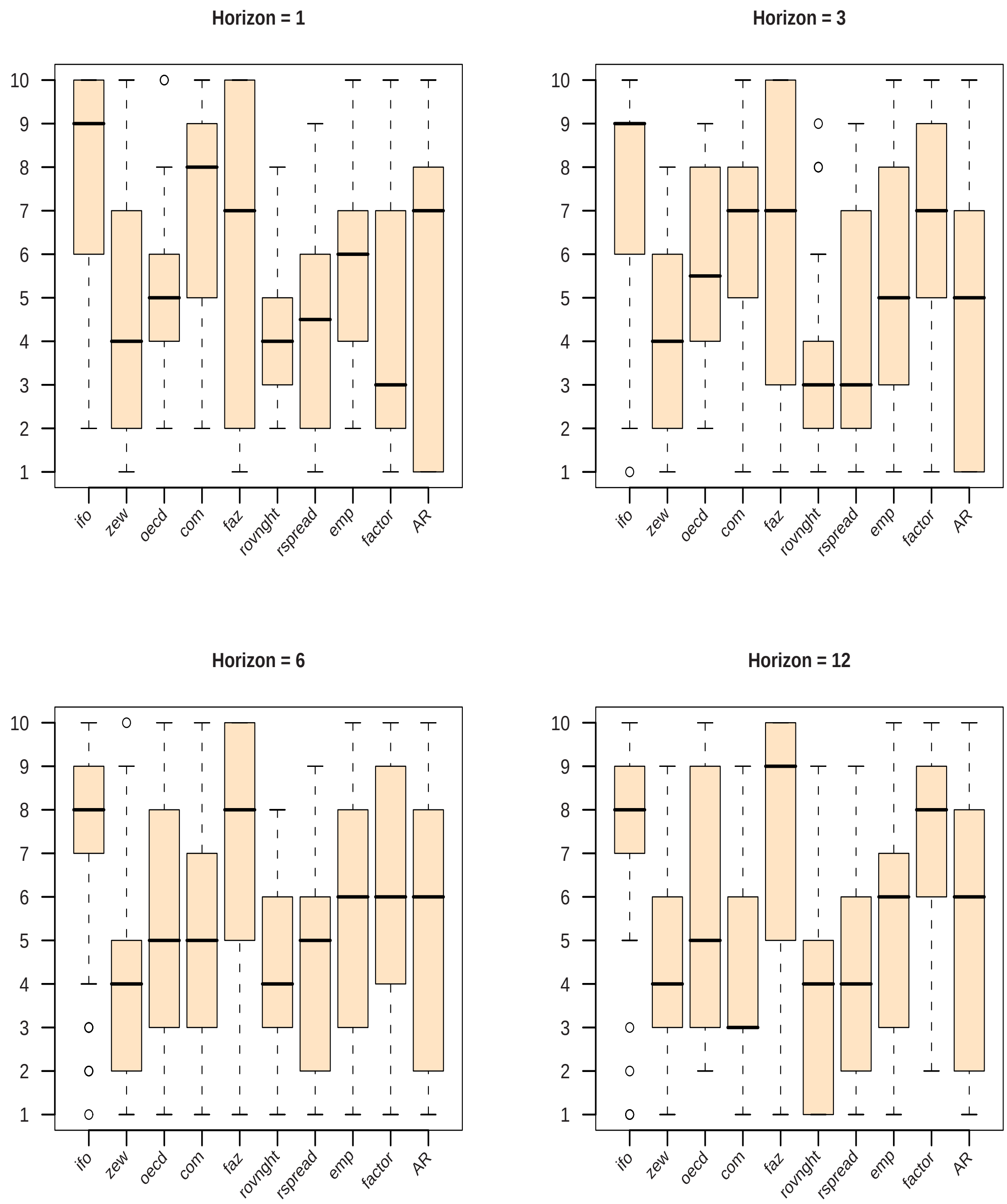


Table 7: Corrected Diebold-Mariano Tests: Exact Yearly Growth Rates

					$h = 1$					
	ifo	zew	com	oecd	faz	rovvnght	rspread	emp	factor	AR
ifo	0	17	25	26	38	22	19	14	49	40
zew	0	0	2	4	12	3	2	0	18	16
com	0	0	0	0	4	2	2	1	17	0
oecd	0	3	4	0	20	4	4	0	23	14
faz	0	0	0	0	0	3	4	2	14	6
rovnght	2	3	10	2	20	0	5	2	23	25
rspread	0	4	10	2	23	7	0	0	21	28
emp	0	4	13	3	36	15	4	0	31	30
factor	0	0	2	0	0	1	2	1	0	6
AR	4	4	6	6	8	8	8	4	12	0
					$h = 3$					
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	29	26	26	36	25	27	16	40	36
zew	2	0	2	1	18	2	4	2	27	1
com	2	2	0	2	7	1	6	4	9	0
oecd	2	12	7	0	29	5	8	6	31	14
faz	0	4	0	0	0	0	4	1	7	8
rovnght	2	2	6	0	12	0	6	4	17	12
rspread	0	6	4	2	12	0	0	1	25	20
emp	0	5	9	5	24	11	11	0	40	24
factor	0	0	0	0	7	1	4	3	0	9
AR	2	0	4	2	10	6	6	7	14	0
					$h = 6$					
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	5	2	8	13	7	9	7	23	8
zew	3	0	5	9	12	3	4	4	20	1
com	0	2	0	0	11	0	5	3	5	0
oecd	11	11	11	0	19	10	13	11	33	8
faz	0	1	8	0	0	4	5	4	7	2
rovnght	0	2	5	0	15	0	6	4	12	0
rspread	7	6	2	9	8	0	0	0	19	19
emp	10	12	7	10	17	5	8	0	31	24
factor	0	0	0	0	2	2	6	4	0	3
AR	6	7	7	8	14	6	8	8	23	0
					$h = 12$					
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	7	6	5	10	14	9	6	12	0
zew	15	0	12	15	15	8	25	9	23	7
com	0	2	0	2	8	6	6	4	6	0
oecd	12	13	13	0	15	11	15	14	19	8
faz	15	9	16	12	0	15	17	18	18	15
rovnght	3	6	7	3	11	0	6	6	7	0
rspread	5	6	2	7	9	0	0	1	8	2
emp	5	9	1	14	11	8	10	0	16	4
factor	0	0	3	0	5	4	6	4	0	0
AR	12	16	6	17	12	11	17	10	18	0

Table 8: Corrected Diebold-Mariano Tests: Appr. Yearly Growth Rates

	ifo	zew	com	oecd	faz	rovvnght	rspread	emp	factor	AR
						$h = 1$				
ifo	0	21	27	26	41	26	21	21	50	44
zew	0	0	2	3	13	3	2	0	20	16
com	0	2	0	0	6	2	3	2	17	4
oecd	0	3	4	0	16	4	4	1	28	10
faz	0	0	0	0	0	2	2	2	12	8
rovnght	2	2	14	2	18	0	3	2	24	23
rspread	0	3	12	1	22	11	0	2	22	28
emp	0	4	15	3	34	13	4	0	32	30
factor	0	0	3	0	0	2	2	1	0	6
AR	4	4	7	4	7	8	8	4	12	0
						$h = 3$				
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	28	26	25	33	22	26	16	42	36
zew	1	0	1	1	15	2	3	2	24	1
com	2	3	0	2	8	2	6	4	8	0
oecd	2	10	7	0	25	3	8	6	32	15
faz	0	4	0	0	0	0	5	1	11	8
rovnght	2	2	5	0	10	0	6	4	13	11
rspread	0	7	6	1	9	0	0	2	24	19
emp	0	7	9	4	25	10	10	0	38	25
factor	0	0	0	0	2	0	3	2	0	9
AR	2	0	4	2	9	6	6	7	14	0
						$h = 6$				
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	6	2	8	13	9	9	7	23	11
zew	3	0	4	9	11	2	2	1	18	3
com	0	2	0	0	9	0	4	2	5	0
oecd	11	12	13	0	19	9	13	11	34	8
faz	0	1	8	0	0	4	3	3	12	3
rovnght	0	2	4	0	12	0	6	4	12	0
rspread	6	6	2	10	8	0	0	0	19	19
emp	10	11	7	9	15	6	7	0	31	22
factor	0	0	0	0	1	0	5	4	0	8
AR	6	5	7	6	14	5	7	6	22	0
						$h = 12$				
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	7	6	6	10	15	9	6	12	0
zew	16	0	10	14	15	8	24	8	24	7
com	0	2	0	2	8	5	6	4	6	0
oecd	12	13	13	0	15	11	15	14	19	8
faz	15	10	17	12	0	16	17	18	19	15
rovnght	2	6	7	3	11	0	6	6	7	0
rspread	5	6	2	7	9	1	0	1	8	2
emp	5	8	1	15	11	9	9	0	16	4
factor	0	0	4	0	4	5	5	4	0	0
AR	12	15	6	16	12	12	16	10	18	0

Table 9: Corrected Diebold-Mariano Tests: Exact Monthly Growth Rates

	ifo	zew	com	oecd	faz	rovvnght	rspread	emp	factor	AR
						$h = 1$				
ifo	0	28	6	21	11	17	22	23	34	22
zew	6	0	0	11	7	6	7	2	10	10
com	6	1	0	4	8	6	5	4	22	10
oecd	6	0	1	0	8	4	4	4	5	10
faz	1	4	5	3	0	4	4	3	14	15
rovnght	4	0	0	2	5	0	5	2	20	10
rspread	4	0	4	3	6	4	0	3	20	12
emp	10	14	0	2	6	5	20	0	20	14
factor	3	0	0	0	4	0	0	0	0	10
AR	4	18	4	4	8	16	24	14	34	0
						$h = 3$				
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	6	11	5	10	10	13	9	10	0
zew	0	0	0	1	7	5	9	0	2	0
com	5	4	0	4	6	4	10	6	2	0
oecd	4	6	4	0	6	5	10	5	6	4
faz	2	5	11	1	0	4	9	11	9	3
rovnght	1	2	4	2	7	0	14	4	4	0
rspread	2	1	6	4	8	1	0	8	5	0
emp	1	0	2	1	6	1	9	0	2	0
factor	3	3	7	6	7	6	8	4	0	1
AR	4	6	9	4	12	10	17	12	6	0
						$h = 6$				
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	13	13	9	2	11	19	13	4	5
zew	4	0	9	4	4	4	6	3	3	2
com	4	5	0	2	1	5	5	6	2	2
oecd	2	5	10	0	1	7	10	7	3	0
faz	9	11	10	5	0	10	11	11	9	7
rovnght	0	0	6	2	2	0	7	6	1	2
rspread	2	2	4	4	2	4	0	3	1	2
emp	0	0	7	5	1	2	6	0	2	4
factor	1	1	13	8	2	5	6	6	0	3
AR	5	3	2	7	5	2	7	8	3	0
						$h = 12$				
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	9	5	5	2	8	8	6	5	12
zew	6	0	3	4	6	3	4	4	2	14
com	6	4	0	4	4	6	6	6	5	13
oecd	8	8	6	0	6	9	11	10	5	14
faz	7	12	8	15	0	8	8	10	10	20
rovnght	4	0	2	2	5	0	4	3	4	11
rspread	6	0	2	3	6	1	0	5	0	10
emp	6	2	4	4	4	1	8	0	2	11
factor	9	10	9	4	10	10	11	10	0	20
AR	2	0	0	0	2	2	1	3	0	0

Table 10: Corrected Diebold-Mariano Tests: Appr. Monthly Growth Rates

					$h = 1$					
	ifo	zew	com	oecd	faz	rovvnght	rspread	emp	factor	AR
ifo	0	21	27	26	41	26	21	21	50	44
zew	0	0	2	3	13	3	2	0	20	16
com	0	2	0	0	6	2	3	2	17	4
oecd	0	3	4	0	16	4	4	1	28	10
faz	0	0	0	0	0	2	2	2	12	8
rovnght	2	2	14	2	18	0	3	2	24	23
rspread	0	3	12	1	22	11	0	2	22	28
emp	0	4	15	3	34	13	4	0	32	30
factor	0	0	3	0	0	2	2	1	0	6
AR	4	4	7	4	7	8	8	4	12	0
					$h = 3$					
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	28	26	25	33	22	26	16	42	36
zew	1	0	1	1	15	2	3	2	24	1
com	2	3	0	2	8	2	6	4	8	0
oecd	2	10	7	0	25	3	8	6	32	15
faz	0	4	0	0	0	0	5	1	11	8
rovnght	2	2	5	0	10	0	6	4	13	11
rspread	0	7	6	1	9	0	0	2	24	19
emp	0	7	9	4	25	10	10	0	38	25
factor	0	0	0	0	2	0	3	2	0	9
AR	2	0	4	2	9	6	6	7	14	0
					$h = 6$					
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	6	2	8	13	9	9	7	23	11
zew	3	0	4	9	11	2	2	1	18	3
com	0	2	0	0	9	0	4	2	5	0
oecd	11	12	13	0	19	9	13	11	34	8
faz	0	1	8	0	0	4	3	3	12	3
rovnght	0	2	4	0	12	0	6	4	12	0
rspread	6	6	2	10	8	0	0	0	19	19
emp	10	11	7	9	15	6	7	0	31	22
factor	0	0	0	0	1	0	5	4	0	8
AR	6	5	7	6	14	5	7	6	22	0
					$h = 12$					
	ifo	zew	com	oecd	faz	rovnght	rspread	emp	factor	AR
ifo	0	7	6	6	10	15	9	6	12	0
zew	16	0	10	14	15	8	24	8	24	7
com	0	2	0	2	8	5	6	4	6	0
oecd	12	13	13	0	15	11	15	14	19	8
faz	15	10	17	12	0	16	17	18	19	15
rovnght	2	6	7	3	11	0	6	6	7	0
rspread	5	6	2	7	9	1	0	1	8	2
emp	5	8	1	15	11	9	9	0	16	4
factor	0	0	4	0	4	5	5	4	0	0
AR	12	15	6	16	12	12	16	10	18	0

Ifo Working Papers

No. 56 Grundig, B., Why is the share of women willing to work in East Germany larger than in West Germany? A logit model of extensive labour supply decision, February 2008.

No. 55 Henzel, S., Learning Trend Inflation – Can Signal Extraction Explain Survey Forecasts?, February 2008.

No. 54 Sinn, H.-W., Das grüne Paradoxon: Warum man das Angebot bei der Klimapolitik nicht vergessen darf, Januar 2008.

No. 53 Schwerdt, G. and J. Turunen, Changes in Human Capital: Implications for Productivity Growth in the Euro Area, December 2007.

No. 52 Berlemann, M. und G. Vogt, Kurzfristige Wachstumseffekte von Naturkatastrophen – Eine empirische Analyse der Flutkatastrophe vom August 2002 in Sachsen, November 2007.

No. 51 Huck, S. and G.K. Lünser, Group Reputations – An Experimental Foray, November 2007.

No. 50 Meier, V. and G. Schütz, The Economics of Tracking and Non-Tracking, October 2007.

No. 49 Buettner, T. and A. Ebertz, Quality of Life in the Regions – Results for German Counties, September 2007.

No. 48 Mayr, J. and D. Ulbricht, VAR Model Averaging for Multi-Step Forecasting, August 2007.

No. 47 Becker, S.O. and K. Wohlrabe, Micro Data at the Ifo Institute for Economic Research – The "Ifo Business Survey", Usage and Access, August 2007.

No. 46 Hülsewig, O., J. Mayr and S. Sorbe, Assessing the Forecast Properties of the CESifo World Economic Climate Indicator: Evidence for the Euro Area, May 2007.

No. 45 Buettner, T., Reform der Gemeindefinanzen, April 2007.

No. 44 Abberger, K., S.O. Becker, B. Hofmann und K. Wohlrabe, Mikrodaten im ifo Institut – Bestand, Verwendung und Zugang, März 2007.

No. 43 Jäckle, R., Health and Wages. Panel data estimates considering selection and endogeneity, March 2007.

No. 42 Mayr, J. and D. Ulbricht, Log versus Level in VAR Forecasting: 16 Million Empirical Answers – Expect the Unexpected, February 2007.

No. 41 Oberndorfer, U., D. Ulbricht and J. Ketterer, Lost in Transmission? Stock Market Impacts of the 2006 European Gas Crisis, February 2007.

No. 40 Abberger, K., Forecasting Quarter-on-Quarter Changes of German GDP with Monthly Business Tendency Survey Results, January 2007.

No. 39 Batchelor, R., Forecaster Behaviour and Bias in Macroeconomic Forecasts, January 2007.

No. 38 Sülzle, K., Innovation and Adoption of Electronic Business Technologies, December 2006.

No. 37 Overesch, M. and G. Wamser, German Inbound Investment, Corporate Tax Planning, and Thin-Capitalization Rules – A Difference-in-Differences Approach, December 2006.

No. 36 Kempkes, G. and C. Pohl, The Efficiency of German Universities – Some Evidence from Non-Parametric and Parametric Methods, October 2006.

No. 35 Kuhlmann, A., German Productivity – A Reassessment via the New Ifo Productivity Database, October 2006.

No. 34 Kuhlmann, A., What is the X-Factor in the German Electricity Industry?, September 2006.

No. 33 Temple, J. and L. Wößmann, Dualism and cross-country growth regressions, August 2006.

No. 32 Baumann, F., V. Meier and M. Werding, Transferable Provisions in Individual Health Insurance Contracts, July 2006.

No. 31 Abberger, K., Qualitative Business Surveys in Manufacturing and Industrial Production – What can be Learned from Industry Branch Results?, May 2006.

No. 30 Ruschinski, M., Investigating the Cyclical Properties of World Trade, May 2006.

No. 29 Holzner, Chr., V. Meier and M. Werding, Time Limits in a Two-tier Unemployment Benefit Scheme under Involuntary Unemployment, April 2006.