

Freedom of Choice in Macroeconomic Forecasting*

Nikolay Robinzonov

Institut für Statistik

Ludwig-Maximilians-Universität

Ludwigstraße 33

80539 Munich

Nikolay.Robinzonov@stat.uni-muenchen.de

Klaus Wohlrabe[†]

Ifo Institute for Economic Research

Poschingerstrasse 5

81679 Munich

wohlrabe@ifo.de

March 1, 2009

Abstract

Different studies provide a surprisingly 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 two possible representations of industrial production. Further on we consider a large variety of time-varying specifications. In a horse race with nine leading indicators plus an AR benchmark model 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. Nevertheless, the freedom of choice can be useful to identify robust leading indicators.

JEL: C52, C53, E37

Keywords: forecasting competition, leading indicators, model selection

*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 their 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? Which model and which leading indicator do I employ? 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) forecasts, or those based on other indicators. A practitioner looking at the empirical literature finds 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 some 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 many different options within the forecasting process. Among these decisions probably the most important point is the employed time series model and its specification.² 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 the economic environment and the objective of the exercise.

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 investigate the forecasting performance of non-linear time series models³ and

¹More details can be found in the literature section.

²Elliot and Timmermann (2008) review almost all issues concerning economic forecasts. In an empirical application the authors investigate the performance of several time series models by forecasting inflation and stock returns.

³See Clements, Franses, and Swanson (2004) for a literature overview, Teräsvirta, van

large scale factor models.⁴ Besides the model comparison a focus has been put on assessments of the forecast performance of a leading indicator. Based on the assumption that an indicator and a 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 other possible model specifications.

Consequently we ask: Does the forecast performance of a leading indicator depend on the forecasting setting? The answer turns out to be "yes", which is not surprising. We conduct a comprehensive study by covering almost all commonly used linear forecasting techniques. This is made by forecasting German industrial production (IP) growth rates with nine leading indicators. We demonstrate how the assessment of the forecasting properties may differ between different forecasting settings. Our results are in line with previous papers on forecasting German industrial production, which differ in the assessment of the used indicators. Our contribution is to illustrate various (often similar) specifications and the corresponding assessment of indicators in one paper. We seek to investigate whether it is possible to identify a robust indicator that proves to have good forecasting properties across settings.

We use the seasonally adjusted monthly industrial production for Germany. As we focus on stationary time series models, we forecast two stationary representations of the target variable. We calculate the exact monthly and yearly growth rates. We consider two different time series models that can be considered the 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. Finally we end up with 50 forecasting settings for each indicator.

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.

⁴See Stock and Watson (2002), Forni, Hallin, Lippi, and Reichlin (2003), Dreger and Schumacher (2005), Schumacher (2007), and Eickmeier and Ziegler (2008) among others.

From a theoretical point of view it would be best practice to choose the "best" specification from those that are available. Although great care is generally taken in designing a specific forecasting model, the true forecast uncertainty is often underestimated because various sources of forecasting errors, like parameter and model uncertainties, are not taken into account properly. However, information not selected for forecasting might be useful. Granger and Jeon (2004b) coined the term "thick modeling" for using many alternative specifications of similar quality. It is well-known in the forecasting literature that the combination of forecasts is often a better procedure than using the individually best forecast. We seek to investigate if we can improve the best single forecast by a combination across models, indicators and observations windows.

The paper is structured as follows. In section 2 we review the existing literature for forecasting German industrial production. Each contribution uses a different transformation of the target variable and forecasting model. We show that the assessment of indicators is different in every reviewed paper. Then we illustrate the freedom of choice in macroeconomic forecasting, i.e. the very similar possible specifications in using linear forecasting techniques. Section 4 illustrates the data issues and provides details of the leading indicators used. The empirical results of the comprehensive forecasting competition are presented in section 5. In section 6 we demonstrate how the best single forecast performance can be improved by combinations of individual forecasts. Then we discuss our results and conclude.

2 Review of the literature for forecasting industrial production in Germany

As we focus on industrial production (IP) for Germany, we review this particular strand of recent literature. Table 1 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 differ for various aspects. Although all papers use industrial production as the reference series, they are not identical. Besides the article by Fritsche and Stephan (2002), who start 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.⁵ All papers apply variations of a VAR model and do a recursive forecasting exercise. As a benchmark model they use an AR model. 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 producers 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 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 it is only statistically significant for a horizon of 12 months. 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

⁵It is interesting to note that the authors relate their own results to the previous papers. In a strict sense the articles are not comparable.

difficult, as the target time series are not identical. The assessment strongly depends on the definition of time series and forecasting settings.

Table 1: Literature review of forecasting German IP

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-Mariano 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

3 Methodology - The freedom of choice

In the introduction we presented some of the 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 that an investigator is confronted with. Table 2 displays the most important options. Ultimately, we end up with 50 possible forecasting settings for a specific time series, which illustrates the freedom of choice in macroeconomic forecasting.

Table 2: Data and Model Considerations

Data	Estimation Window	Forecasting Approach	Model	Restrictions	Selection Criterion
monthly exact	rolling	direct	ARX(p, r)	yes	AIC
yearly exact	recursive	indirect	VAR(p)	no	BIC OSC

3.1 Estimation window: 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 2006:09 minus the forecast horizon. We forecast 1, 3, 6 and 12 months ahead for each approach. In the end we generated 45 forecasts for each horizon. For the rolling forecast, the data vintage consists of 120 observations (1992:01 - 2001:12) which is moved forward in time.

3.2 Forecasting approach: Direct vs. indirect

Forecasts can be generated in two different ways: iterated (indirect or "plug-in") and directly. The iterated forecasts entail estimating an autoregression

and then iterating upon that autoregression to obtain the multiperiod forecast.⁶ The direct forecast entails regressing a multiperiod-ahead value of the dependent variable on current and past values of the variable.⁷ For example, forecasting the industrial production directly twelve months from now might entail the regression of the IP, twelve months hence, against 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. Using a large data set of 170 US monthly macroeconomic time series, Marcellino, Stock, and Watson (2006) demonstrate that iterated forecasts typically outperform the direct forecasts, particularly 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.

Ultimately 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.

3.3 Information set: 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 period being forecast is employed ("perfect foresight"). 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)).

⁶Using formulae: y_t is regressed on lagged values of y_t and the indicator x_t . The forecast y_{t+h} is obtained by calculating y_{t+1} through y_{t+h} . See also equation (1) in section 3.4.

⁷In this case, y_{t+h} is regressed on lagged values of y_t and the indicator x_t . The forecast y_{t+h} is obtained directly without calculating $y_{t+1} \dots y_{t+h-1}$. See also equation (2) in section 3.4.

3.4 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.

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

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

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

$$y_{t+h} = \beta + \sum_{i=1}^p \delta_i y_{t+1-i} + \sum_{j=1}^r \gamma_j x_{t+1-j} + \varepsilon_{t+h}. \quad (2)$$

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 (1) and (2) to the bivariate case. We consider the following VAR(p) model

$$\mathbf{y}_t = \boldsymbol{\alpha} + \sum_{i=1}^p \mathbf{A}_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t \quad (3)$$

where \mathbf{y}_t is now a 2×1 vector containing the IP and the indicator variable, the \mathbf{A}_i are fixed (2×2) coefficients matrices, $\boldsymbol{\alpha}$ is a fixed 2×1 vector of intercept terms and finally $\boldsymbol{\varepsilon}_t$ 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.⁸ 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.

3.5 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 may lead to the curse of dimensionality or overparameterization problems (overfitting). Due to insufficient degrees of freedom, the model parameters are then imprecisely estimated, yielding large standard errors and high estimation uncertainty.⁹ 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 a curse of dimensionality and overfitting the regression.

We employ two of the most popular selection criteria: the Akaike information criterion (AIC) and the Bayesian 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.¹⁰

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

⁸See footnote 2 for references.

⁹This 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).

¹⁰See e.g. Mills and Prasad (1992). Granger and Jeon (2004a) 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.

model for each indicator is the one with the lowest mean squared forecast error over the respective forecast horizon. In the empirical application we split the sample into three parts. The first part is the estimation sample, then the evaluation sample (3 years), where we choose the model with the lowest root mean squared error (RMSE) over the forecasting horizons. With the selected model we calculate the forecasts. We move these windows forwards in our out-of-sample forecast exercise. In simple words, under the OSC we assume that the model performed well in the past and will also do so for the current situation.¹¹

Inoue and Kilian (2006) investigate all three criteria for choosing forecasting models.¹² 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 RMSE yields inconsistent results and have a probability greater than zero of choosing a model which does not have the best forecasting performance, while the BIC is consistent for nested models.¹³

3.6 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.

¹¹See Swanson and White (1997) for a systematic investigation concerning out-of-sample model selection.

¹²The authors consider only nested models.

¹³Elliot and Timmermann (2008) state that consistency is not the most important criterion in forecasting.

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

4 The data

4.1 German Industrial Production

The target variable in our case study is industrial production (IP) for Germany from 1991:01 to 2006:09. We do not use data before 1991 in order to circumvent any structural breaks in the data due to 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 Deutsche Bundesbank.¹⁴

In our case study we only use stationary time series models. Therefore we focus on stationary representations (interpretations) of the (trending) German industrial production. In general one can calculate exact and approximate (log differences) monthly and yearly growth rates. One can interpret the monthly growth rates as rather short-run dynamics whereas the yearly growth rates refer to longer trends over time. All these transformations have been used in the literature (see section 2). In our paper we use the exact calculation of growth rates.¹⁵ Figure 1 plots the monthly and yearly growth rate. The left graph shows the exact yearly growth rates, which exhibit a clear periodical pattern. The monthly growth rates (right panel) display an

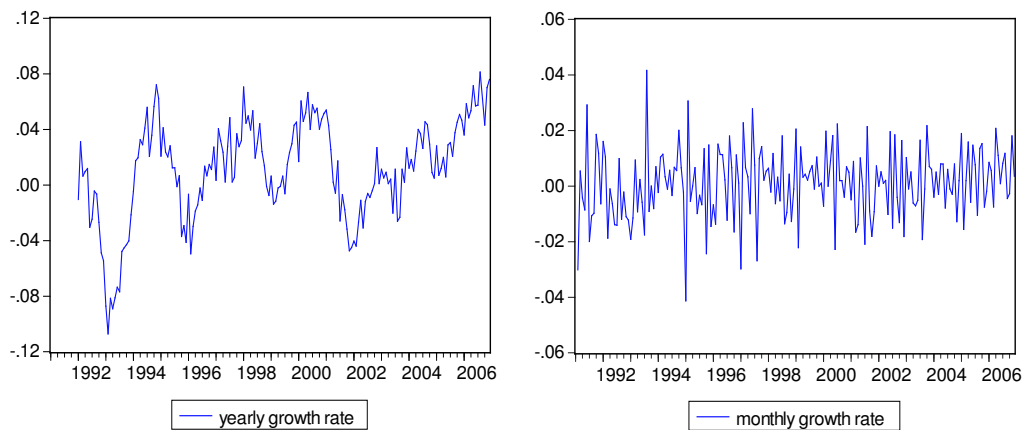
¹⁴Series USNA01.

¹⁵Still, it could be interesting if the differences between exactly and approximately calculated growth rates lead to different conclusions about the assessment of indicators. See Robinsonov and Wohlrabe (2008) for a comparison.

erratic pattern and seem to be harder to forecast.¹⁶

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 or first differences is forecasted.¹⁷ Second, the choice of a specific transformation is rarely justified in the literature. In our literature review no article motivates the employed data transformation.¹⁸ 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.

Figure 1: Representations of Industrial Production in Germany



4.2 Leading indicators

In order to illustrate the diversity of forecasting outcomes, we conduct our forecasting exercise with nine leading indicators displayed in Table 3. 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.

¹⁶There are further possible transformations of the target variable; see Marcellino (2006) for examples and references.

¹⁷One could transform the forecast back into the original level series and judge this forecast accuracy but this is not usually done.

¹⁸A natural choice in many papers is the yearly growth rate transformation.

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 as to the expected economic development in Germany in six months; see Hübner 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 (manufacturing, % balance), Ifo Business Climate Indicator (manufacturing, % balance), spread of interest rates (% annual rate), total new orders (manufacturing), finished goods stocks (manufacturing, level) and Export order books (manufacturing, level).

In addition to survey and composite indicators we take some financial indicators as a possible predictors. Since the seminal paper by Estrella and

Table 3: Leading Indicators

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

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 that 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 definitions see Table 3.

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 2006.¹⁹ 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.

5 Empirical results

In our case study we forecast two representations of German IP 1, 3, 6 and 12 months ahead. The initial forecast date is 2002:01 and the final forecast date

¹⁹The estimated factor was kindly provided by Christian Schumacher and is based on the paper Marcellino and Schumacher (2007).

is 2006:09 minus the forecast horizon. We employ the ARX and the VAR models for each of the time series. The lag selection is made via the common information criteria AIC, BIC and OSC (with and without restrictions²⁰) and the ARX-specific approaches ex-post and ex-ante. For the OSC the evaluation sample consists of four years. For the initial forecast 2002:01 we choose the best model via OSC over the horizon 1998:01-2001:12. We move this window forward through the forecasting exercise. For every specification we conduct the direct and indirect forecasting techniques and finally extend these to both time varying schemes: rolling and recursive forecasting. In combination these settings sum up to 50 forecasting specifications for each horizon.

With 9 indicators and 2 time series to be forecasted we have 18 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 model 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 three 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. Given the large information set from the various forecasting settings we show how forecast combinations may improve the forecast performance. Then we discuss some issues on assessing leading indicators.

5.1 General remarks

Table 4 tabulates the descriptive statistics of correlations between the forecasted and the actual values. The mean correlation describes the average correlations over all indicator model combinations. Considering monthly growth rates we find that only for one month ahead do the majority of models/indicators combinations provide information content. For horizons beyond three months, the average forecasts are purely noise (average correlations of almost zero) although there is a model/indicator combination with a relative positive correlations around 0.4. Considering yearly growth rates the correlations are higher. For horizons up to three months ahead the correlations distribution is tight and with a mean of 0.78. This indicates that the different model/indicator combinations yield similar results. As expected

²⁰Due to very large possible lag combinations we abstain from restricted forecasts in the OSC case.

the correlation mean decreases with an increasing forecasting horizon. Nevertheless for $h = 12$ one can detect a maximum correlation of 0.90 and thus a suitable model for long-term forecasting.

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. The main reason is that in a (iterative) VAR setting both variables are forecasted with their own past values and the other variable. This can introduce higher forecasting errors just because the leading indicator is forecasted poorly.

Furthermore in about 70% of the cases the rolling scheme produces lower RMSE than the corresponding recursive scheme. The OSC criterion delivers by far lower RMSEs than the statistical selection criteria AIC and BIC. Finally, the indirect approach outperforms the direct approach. These results are similar to previous findings in the literature.

Table 4: Correlation of forecasts with actual values

Horizon	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Yearly growth rates						
1	0.67	0.76	0.78	0.78	0.81	0.89
3	0.23	0.58	0.65	0.63	0.69	0.84
6	-0.21	0.22	0.35	0.34	0.47	0.84
12	-0.82	-0.45	-0.18	-0.08	0.24	0.90
Monthly growth rates						
1	-0.28	0.47	0.53	0.42	0.56	0.67
3	-0.34	0.01	0.18	0.14	0.29	0.51
6	-0.40	-0.08	0.00	-0.01	0.08	0.43
12	-0.45	-0.12	0.01	0.01	0.13	0.40

5.2 The winners

For each setting/indicator combination we calculate the Root Mean Squared Error (RMSE). Tables 5 and 6 present the best indicator for each model specification chosen by the lowest RMSE. The indicator in bold face displays for each forecast horizon the minimum RMSE over all model/indicator combinations.

For the exact yearly growth rates, the AR benchmark model, the FAZ indicator (*faz*) and the *factor* are the dominant winners. But also the fi-

nancial indicators (*rspread* and *rovnght*), the Early Bird indicator (*com*) and the OECD indicator (*oecd*) provide the lowest RMSE in some specifications. In the short run the unrestricted indirect ARX with the *factor* provides the lowest RMSE. For the other horizons the *faz* indicator performs best within the ARX framework. Using the direct approach, the AR benchmark can hardly be outperformed by an indicator. Furthermore, we can state that there are only few differences in the results between the rolling and the recursive forecasting scheme. In the majority of all settings we find for both schemes almost always the same indicator with the lowest RMSE.

If we look at the exact monthly growth rates (Table 4), we find a heterogeneous picture. Financial indicators (*rspread* and *rovnght*) play a more dominant role. This is not surprising as monthly growth rates display short run dynamics of an economy. 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. Again, hardly any indicator can outperform the AR benchmark within the direct ARX settings. Comparing the rolling and recursive setting, we find many cases where the winner is different between the estimation windows.

Table 5: Forecast Competition - Yearly Growth Rates - Best Indicator

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

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

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

5.3 The ordinal ranking

Tabulating only the winners for each setting does not yield an impression of the variance of assessment over different forecasting settings. For each forecasting setting we rank the nine indicators plus the AR benchmark due to the RMSE criterion. The lowest model/indicator combination with the lowest RMSE is ranked first and so on. For each time series, horizon and indicator we draw the boxplot over all 50 forecasting settings.²¹ Figures 2 and 3 display the corresponding ranking boxplots. The boxplots can be read as follows. The ordinate refers to the possible ranks from 1st to 10th place. The colored box for each indicator represents the interquartile range of the rank distribution. The lower end is the first and the upper end is the third quartile, respectively. The bold line indicates the median of all 50 ranking positions. The dashed lines outline the occurrence of outliers outside the interquartile range.²² The circles refer to extreme outliers, i.e. ranks that occur only once. A robust ranking of a specific indicator across forecasting settings is illustrated by a small box with short dashed lines.

For the exact yearly growth rates (Figure 2), we see that the *factor* indicator is relatively robust across settings and horizons.²³ 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 horizons (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.²⁴ 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* indicator performs well because it is constructed to lead industrial production. For $h = 12$ we find a very heterogeneous picture, where no indicator is clearly

²¹Consider the indirect unrestricted ARX(p, r) setting with model selection due to the AIC criterion. For each indicator we calculate the corresponding average RMSE. Now we can rank the indicators in this model class due to the RMSE criterion. We repeat this procedure for each forecasting setting. Finally we end up with 50 ranks for each indicator. This distribution of ranks is summarized in the boxplots.

²²Its maximum length is 1.5 times the interquartile range.

²³The corresponding box is small and the median is on the first place.

²⁴It contains also survey information from the construction, wholesale and retail sectors. 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.

dominant. The *com* indicator is, on average, the best for forecasting more than six months ahead.

In the case of the monthly growth rates, the ranking variance is much more pronounced (Figure 3). It is always possible by comparing two specific indicators to find a forecasting setting where one indicator is better than another and vice versa. There is no dominant indicator across forecasting settings and horizons. For illustration purposes let us consider the forecast three months ahead of the exact monthly growth rates and compare two indicators: *ifo* and *zew*. Employing an iterative VAR model with a rolling scheme and the OSC criterion *zew* outperforms all indicators while *ifo* is evaluated as the worst one. For the same horizon and times series under the iterative ARX model with a recursive scheme and OSC criterion, the *zew* indicator strongly deteriorates to the ninth place while *ifo* becomes the winner. The same large quality magnitude can be found between *faz* and *com* for the exact yearly growth rates as well.

5.4 Forecasting combinations

In the previous section we have generated many multiple forecasts for the same variable. Decision makers are often faced with such multiple forecasts which often reflect differences in forecasters subjective judgements due to heterogeneity in their information sets. One can ask: Should a single dominant forecast be identified or should a combination of the underlying forecasts be used to produce a pooled summary measure? Forecast combinations have been applied successfully in many forecasting areas, see Clemen (1989) for an early review. Recently, Stock and Watson (2004) undertook an extensive study across numerous economic and financial variables using linear model and found that, on average, pooled forecasts outperform prediction from the best single model. There are several reasons for forecast combinations, e.g. individual forecasts may be affected by structural breaks, or are subject to misspecifications. Furthermore forecast combination can be used as a guard against ex ante forecast uncertainty.²⁵ In practice it has often been found that simple combination schemes (as equal weights) do better than more sophisticated rules (as time-varying or non-linear weights), see Bunn (1985), Makridakis and Winkler (1983) or Palm and Nijman (1984), among others.

In this section we explore whether forecast combinations can improve the forecast accuracy of the best single forecasting model for each horizon. We apply two simple combination schemes that proved to be successful in practical applications and are easy to implement. The first one put equal

²⁵See Timmermann (2006) for an comprehensive outline and references.

weights on each forecast and the second assigns weights due to the inverse MSE from the previous period proposed by Stock and Watson (1999). One can combine over three dimensions. First, you can combine over different models, accounting for model uncertainty. Second, you can combine over different estimation periods, accounting for possible structural breaks in the data.²⁶ Third one can combine over information, accounting for possible missing variable problems. This issue is also important as different indicators possibly contain different information. Finally we investigate the aspect of trimming in forecasting combination. In trimming forecasts one discards the models with the worst forecast performance. Granger and Jeon (2004b) recommend trimming five or ten percent of the worst models, whereas Aiolfi and Favero (2005) suggest trimming up to 80% of the the forecasting models.

In the previous section we had 25 forecasting settings for both rolling and recursive estimation schemes. We drop the ex post setting for the combination exercise, as we want to incorporate only information that is available when the forecast is made (ex ante approach). We are left with 40 forecasting settings per indicator and horizon, summing up to 360 possible combinations with nine indicators.

Tables 7 and 8 show the results of the simple forecasting combination exercise. The first line tabulates the lowest RMSE from the best forecasting setting over all 360 possible model-indicator combinations with the corresponding indicator in the second line. We report the RMSE from combining forecasts across models, indicators, and estimation windows in line 3. Then we trimmed the sample by discarding 50% of the worst performing models. The results for the inverse MSE weighting scheme is reported in line 4 and finally the corresponding trimmed combination results. In our empirical example we can considerably outperform the best single forecasting setting only for $h = 12$ for the exact yearly growth rates. For all other horizons for both yearly and monthly growth rates we cannot improve upon the best single forecast.²⁷ In line with the existing literature we find that trimming forecast combinations decreases the RMSE of the combined forecasts. For the monthly growth rates with the equal weighting scheme, we obtain lower RMSEs compared the with inverse MSE weights and vice versa for the yearly growth rates. The success of forecast combinations in many cases in the literature is due to averaging across indicators.²⁸ We employ only 9 indicators. In

²⁶Assenmacher-Wesche and Pesaran (2008) find that averaging across estimation windows is at least as effective as averaging over different models and both complement each other.

²⁷It would be interesting to see if this results holds for more complicated weighting schemes. We leave this for future research.

²⁸For instance, Stock and Watson (2004) employ 43 time series for each of the G7

case of monthly growth rates and inverse MSE weights with 50% forecasting settings trimmed, we obtained quite similar results as the best forecasting setting. Another implication of our results is that it allows us to identify robust indicators across different models. Thus, the *faz* and *factor* are indeed a good choice for forecasting German industrial production.

Table 7: Forecast combinations: Yearly growth rates

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Minimum RMSE	0.100	0.142	0.213	0.401
Indicator	factor	faz	faz	faz
Equal weights (EW)	0.155	0.206	0.362	0.820
Trimmed (50% of EW)	0.135	0.162	0.280	0.612
Inverse MSE weights (IW)	0.169	0.215	0.314	0.369
Trimmed (50% of IW)	0.150	0.190	0.293	0.371

Table 8: Forecast combinations: Monthly growth rates

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Minimum RMSE	0.069	0.101	0.115	0.091
Indicator	factor	rspread	faz	factor
Equal weights (EW)	0.083	0.114	0.121	0.110
Trimmed (50% of EW)	0.074	0.109	0.118	0.104
Inverse MSE weights (IW)	0.106	0.123	0.123	0.133
Trimmed (50% of IW)	0.083	0.103	0.114	0.098

5.5 Discussion

Given our results, provide some comments about assessing an indicator. Generally, the assessment of how "good" or how "bad" an indicator is should be based on a wider basis of information. Conclusions drawn from one or two forecasting settings (out of many more possible ones) can be biased towards a specific indicator. Furthermore, you cannot be sure about the relative performance of your model/indicator combination. Therefore, first, we recommend

economies. They show that in most cases they can improve the best single forecast by forecast combination.

considering at least two model classes and within these model classes many different specifications. From a practical point of view, especially for linear models, it is easy to implement many different forecasting settings. In the case of using only a few forecasting settings, as outlined in our literature review, one is possibly confronted with model uncertainty. Second, we advise considering a wide range of possible indicators. In choosing indicators, 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 that 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 industry sectors that is not commonly known to the public. Third, consider different information sets for forecasting, i.e. employ both a rolling and a recursive estimation scheme.

The combination of many forecasting settings, leading indicators and information sets (possibly) allows us to identify the best, on average, leading indicator that is robust across indicators and forecasting settings.

Given the large generated information set, forecast combination is a natural alternative to improve the forecast accuracy. In our case the simple combination schemes do not in general reduce the RMSE compared to the best single forecasting setting. As already mentioned, an increase in the number of predictors would possibly further improve the performance of combined forecasts. Forecast combination can be used to reduce *ex ante* uncertainty. We showed that trimmed forecast combination come quite close to the best single forecast. Eventually, forecast combinations may possibly allow us to identify a robust indicator.

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 vary across forecasting settings. We illustrate this freedom of choice in a comprehensive case study by forecasting German industrial production with linear time series models. We employ the two forecast 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 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 two representations of industrial production, exact monthly and yearly growth rates. 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 large information set we implemented a simple forecasting combination approach. By averaging all forecasts we were not able to improve upon the best single forecasts.

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 would facilitate the establishment of rich forecasting combinations, which can be better than single forecasts. Furthermore it can reduce ex ante forecasting uncertainty.

References

- AIOLFI, M., AND C. FAVERO (2005): “Model Uncertainty, Thick Modeling and the Predictability of Stock Returns,” *Journal of Forecasting*, 24, 233–254.
- ASSENMACHER-WESCHE, K., AND M. H. PESARAN (2008): “Forecasting the Swiss Economy Using VECX* Models: An Exercise in Forecast Combination Across Models and Observation Windows,” Working Papers 2008-3, Swiss National Bank.
- 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.
- BUNN, D. (1985): “Statistical efficiency in the linear combination of forecasts,” *International Journal of Forecasting*, 1(2), 151–163.

- CHEVILLON, G., AND D. HENDRY (2005): “Non-parametric direct multi-step 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.
- CLEMEN, R. (1989): “Combining forecasts: A review and annotated,” *International Journal of Forecasting*, 5, 559–583.
- 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.
- DENTON, F. (1985): “Data Mining as an Industry,” *The Review of Economics and Statistics*, 67(1), 124–127.
- 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*, 27(1), 237–265.
- ELLIOT, G., AND A. TIMMERMANN (2008): “Economic Forecasting,” *Journal of Economic Literature*, 43(1), 3–56.
- 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.
- FRITSCHKE, 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 (2004a): “Forecasting Performance of Information Criteria with Many Macro Series,” *Journal of Applied Statistics*, 31(10), 1227–1240.
- GRANGER, C., AND Y. JEON (2004b): “Thick modeling,” *Economic Modelling*, 21(2), 323–343.
- 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 R. WINKLER (1983): “Averages of forecasts: Some empirical results,” *Management Science*, 29(9), 987–996.
- MARCELLINO, M. (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.
- PALM, F., AND T. NIJMAN (1984): “Missing Observations in the Dynamic Regression Model,” *Econometrica*, 52(6), 1415–1436.
- ROBINZONOV, N., AND K. WOHLRABE (2008): “Freedom of Choice in Macroeconomic Forecasting: An Illustration with German Industrial Production and Linear Models,” ifo working paper 57, Munich.
- SCHORFHEIDE, F. (2005): “VAR forecasting under misspecification,” *Journal of Econometrics*, 128(1), 99–136.
- SCHUMACHER, C. (2007): “Forecasting German GDP using alternative factor models based on large datasets,” *Journal of Forecasting*, 26(4), 271–302.
- 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.
- (2004): “Combination forecasts of output growth in a seven-country data set,” *Journal of Forecasting*, 23(6), 405–30.
- 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.

Figure 2: Ranking of leading indicators: Yearly growth rates

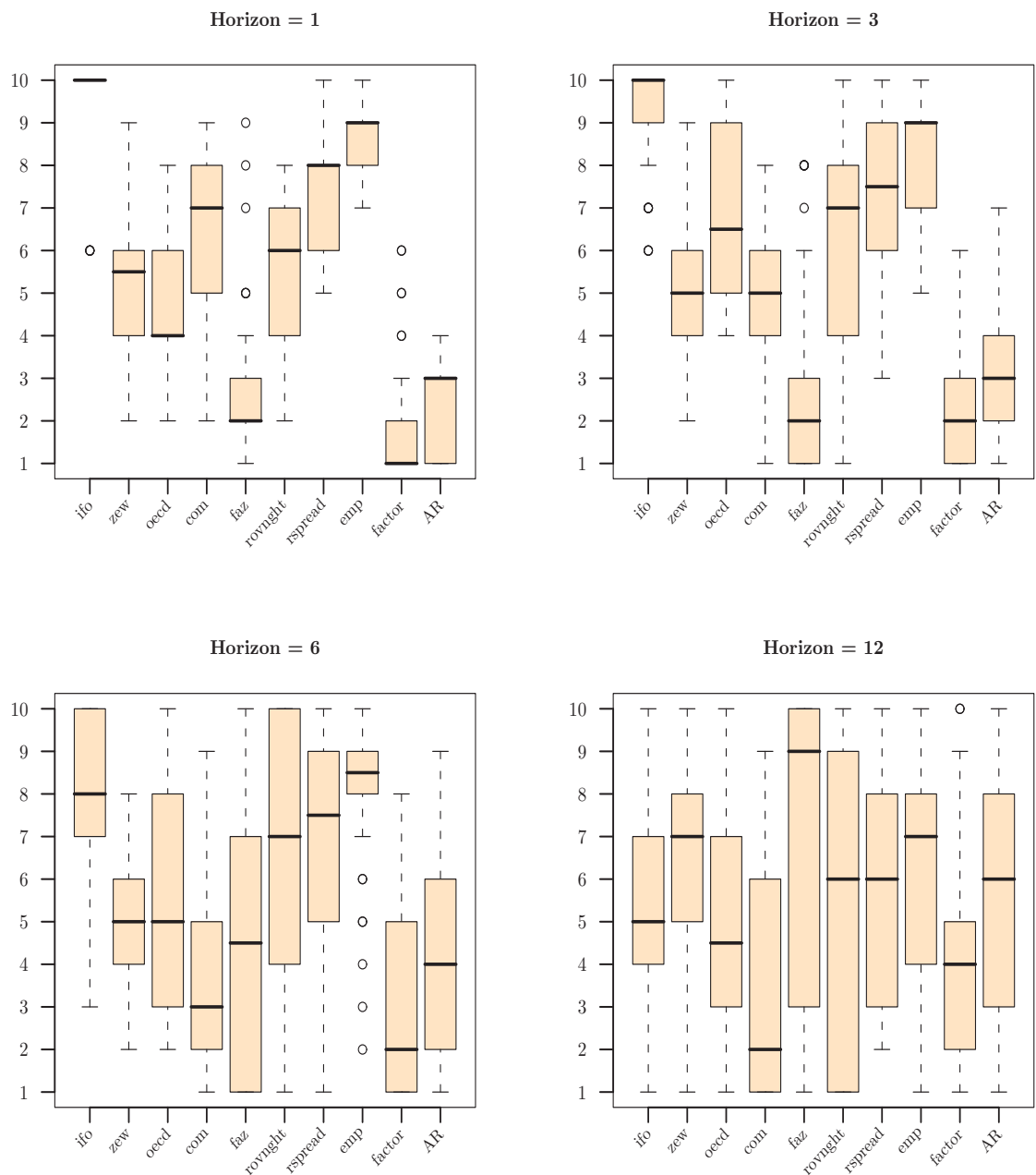


Figure 3: Ranking of leading indicators: Monthly growth rates

