RUHR
ECONOMIC PAPERS

Alexander Schlösser

Forecasting Industrial Production in Germany: The Predictive Power of Leading Indicators

UNIVERSITÄT DUISBURG ESSEN

#838

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #838

Responsible Editor: Volker Clausen

ISSN 1864-4872 (online) – ISBN 978-3-86788-971-1

Ruhr Economic Papers #838

Alexander Schlösser

Forecasting Industrial Production in Germany: The Predictive Power of Leading Indicators

UNIVERSITÄT DUISBURG ESSEN

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/86788971
ISSN 1864-4872 (online)
ISBN 978-3-86788-971-1

Alexander Schlösser[1]

Forecasting Industrial Production in Germany: The Predictive Power of Leading Indicators

Abstract

We investigate the predictive power of several leading indicators in order to forecast industrial production in Germany. In addition, we compare their predictive performance with variables from two competing categories, namely macroeconomic and financial variables. The predictive power within and between these three categories is evaluated by applying Dynamic Model Averaging (DMA) which allows for time-varying coefficients and model change. We find that leading indicators have the largest predictive power. Macroeconomic variables, in contrast, are weak predictors as they are even not able to outperform a benchmark AR model, while financial variables are clearly inferior in terms of their predictive power compared to leading indicators. We show that the best set of predictors, within and between categories, changes over time and depends on the forecast horizon. Furthermore, allowing for time-varying model size is especially crucial after the Great Recession.

JEL-Code: C11, C52, E23, E27

Keywords: Forecasting; industrial production; model averaging; leading indicator; time-varying parameter

January 2020

1 Alexander Schlösser, UDE. – We thank Volker Clausen, Jan Prüser, the participants of the PhD research seminar of the University of Duisburg-Essen as well as the research department of Macroeconomics and Public Finance of the RWI - Leibniz Institute for Economic Research for valuable comments. *– All correspondence to: Alexander Schlösser, University of Duisburg-Essen, Chair of International Economics, Faculty of Economics and Business Administration, Universitätsstraße 12, 45117 Essen, Germany, e-mail: alexander.schloesser@uni-due.de*

1. Introduction

One major branch of economics being continuously well-covered by the media are predictions about the business cycle. The German Council of Economic Experts, several economic research institutes like the RWI, ifo, IfW and many more institutions provide, on a regular basis, forecasts about the future state of the economy. The fact that these forecasts are of fundamental interest to the general public is not surprising as they have immediate effects on current and future decisions. Policymakers are interested in those predictions because they may have to adjust budgets or they want to pass legislation to stabilize the economy. Central banks pay attention to them since economic booms or busts may cause deviations from the inflation target which in turn call for adjustments in the stance of monetary policy. Furthermore, business cycle forecasts are of particular interest to financial market participants to form accurate expectations. Finally, also households are interested in them since changes in the business cycle may influence employment and personal income.

Measures of real activity, like industrial production or even GDP itself, are published with a significant delay. However, policymakers and the private sector have a fundamental interest in detecting economic down- and upturns in a timely manner. This interest has led to the evolution of *leading indicators*. Many institutions spend a large effort to provide economists and practitioners with such indicators. This study evaluates the predictive power of the most prominent leading indicators, namely the ifo business climate, the ZEW indicator, the consumer confidence indicator, the business confidence indicator, the composite leading indicator from the OECD and the euro-coin indicator. But leading indicators are not the only variable category to be considered when forecasting industrial production.[1] Macroeconomic or financial variables may also contain valuable information about the future state of the economy. Therefore, we will also compare the predictive power within as well as across these three categories.

Several studies compare the predictive power of leading indicators and variables from other categories for Germany, e.g. Schumacher and Dreger (2005), Drechsel and Scheufele (2012), Ulbricht et al. (2017) or Heinisch and Scheufele (2018). However, our study dif-

[1] We are aware of the fact that by focusing on industrial production, we are not considering the evolution in the service sector. However, demand for services is less volatile than production since service contracts are usually characterized by a relatively long cancellation period. Therefore, the service sector is less sensitive with respect to the business cycle.

fers from those in several dimensions. Schumacher and Dreger (2005) also compare the predictive power of several leading indicators. But some indicators used in their study have ceased (e.g. the composite leading indicators published by Frankfurter Allgemeine Zeitung or Handelsblatt) and some other promising indicators are not considered (e.g. the composite leading indicator published by OECD or the business confidence indicator from the European Commission). Furthermore, they do not consider macroeconomic and financial variables. Drechsel and Scheufele (2012), in contrast, use a comprehensive set of 120 predictors and focus on the predictive power of leading indicators during the Great Recession in 2008/2009. But due to the large set of predictors and their econometric technique, they do not provide time-varying weights attached to single predictors. We, instead, use data until the recent past which allows us to investigate whether the financial and sovereign debt crisis has caused a persistent change in the weight attached to single predictors and/or in the size of the forecasting model. Another related study is Ulbricht et al. (2017). Even though they also consider leading indicators, their emphasis is on the predictive power of media data. Finally, the predictive power of leading indicators has also been investigated by Heinisch and Scheufele (2018). However, they use mixed-frequency models to predict GDP on a quarterly basis.

One drawback shared by many studies is that they apply model averaging to single predictor models. Furthermore, the models used in these studies do not exhibit the most relevant econometric attributes. That is, a valuable forecasting model should have the following three properties. First, the parameters should be allowed to change over time in order to be able to account for structural breaks. Recursive or rolling forecasting methods can account for time-varying coefficients but only up to a certain extent. Groen et al. (2009) argue that it is better to design a forecasting model that takes time-variation in the parameters explicitly into account. Models with time-varying parameters (TVP) are commonly estimated using the methods described in e.g. Cogley et al. (2005). Many studies like Chan et al. (2012), Ferrara et al. (2015) or Barnett et al. (2014) document the usefulness of TVP-models in forecasting exercises. Second, the number of potential predictors can be large. One strand of the literature has focussed on the development of factor models (see e.g. Stock and Watson (2002)) which are quite successful in forecasting macroeconomic time series. Within this model class, the need of variable selection is eluded by extracting a couple of factors that capture the common movements of all time series. However, many factor models are not able to assess the predictive content of a certain variable category and none of them is able to evaluate the predictive content of a particular variable within a category. Especially the latter is of great interest to many economists and policy makers because it may contain valuable information about the state of the economy. Another strand of the literature has introduced Bayesian Model Averaging (BMA). BMA is attractive since it allows to compute inclusion probabilities for single predictors. In principle, BMA averages over a wide range of models which differ with respect to the predictors and therefore also in size.[2] Third, the model and hence the set of predictors might change over time. For instance, some variables are well suited to predict industrial production in recessions but not in expansions. Ignoring this regime dependence might explain why models considering the whole predictor set at each point in time exhibit poor forecast performance.

This study uses the econometric technique called DMA, developed by Raftery et al. (2010) and introduced to economic applications by Koop and Korobilis (2012). DMA is equipped with all three properties described above. It allows the coefficients of the model to change over time, deals with model selection appropriately and, finally, allows the entire forecasting model to change over time. In so doing, we contribute to the literature in the following ways. We show that the weight attached to all three variable categories is time-varying and that variables from the category of leading indicators receive the largest weight independent from the forecast horizon. Leading indicators exhibit the best quantitative forecasting properties while both, macroeconomic and financial variables, have less predictive power. Our analysis shows that the average model size changes considerably over time and especially during the financial crisis. During this period of considerable economic turmoil, larger models receive a higher weight. Furthermore, there seems to be a positive correlation between model size and forecast horizon as larger models tend to perform relatively better for longer forecasting horizons. Finally, the cumulated absolute forecast error points towards similar forecast performance before and after the Great Recession.

The paper proceeds as follows. Section 2 presents the econometric framework, Section 3 the design of the forecasting experiment, Section 4 the data, Section 5 the empirical results, Section 6 contains robustness checks and, finally, Section 7 concludes.

[2] Hendry and Clements (2004) or Timmermann (2006) show that pooling models can lead to substantial forecast improvements.

2. Dynamic Model Averaging

In order to take parameter and model uncertainty into account, we use the framework described in Koop and Korobilis (2012). Let $k = 1, ..., K$ denote the set of models a professional forecaster wants to consider.[3] Each model can be written in state-space form as

$$\begin{aligned} y_t &= \boldsymbol{x}_t^{(k)} \boldsymbol{\theta}_t^{(k)} + \varepsilon_t^{(k)} \\ \boldsymbol{\theta}_t^{(k)} &= \boldsymbol{\theta}_{t-1}^{(k)} + \eta_t^{(k)} \end{aligned} \tag{1}$$

where $\varepsilon_t^{(k)} \sim N(0, H_t^{(k)})$ and $\eta_t^{(k)} \sim N(\mathbf{0}, \boldsymbol{Q}_t^{(k)})$. Let $M_t \in 1, ..., K$ denote which model applies in each point in time. Since all K models shall be used in forecasting industrial production, we need to calculate the time-varying model probabilities denoted as $P(M_t = k|y^{t-1})$ to average across these models.[4] How $P(M_t = k|y^{t-1})$ is calculated will be discussed below. The beauty of this model is to allow for changes in the marginal effects of each predictor and their corresponding weight. In a first step, we focus on estimating one of these K models.[5] In general, the model could be estimated by applying the Kalman filter and Markov chain Monte Carlo (MCMC) methods. That is, in a first step one would draw $\boldsymbol{\theta}^T$ by applying the conventional Kalman Filter as follows:

Predictions

$$\boldsymbol{\theta}_{t|t-1} = \boldsymbol{\mu} + \boldsymbol{\theta}_{t-1|t-1} \tag{2}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{\Sigma}_{t-1|t-1} + \boldsymbol{Q}_t \tag{3}$$

$$\eta_{t|t-1} = y_t - y_{t|t-1} = y_t - \boldsymbol{x}_t \boldsymbol{\theta}_{t|t-1} \tag{4}$$

$$f_{t|t-1} = \boldsymbol{x}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{x}_t' + H_t \tag{5}$$

Updating

$$\boldsymbol{\theta}_{t|t} = \boldsymbol{\theta}_{t|t-1} + \boldsymbol{K}_t \eta_{t|t-1} \tag{6}$$

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{x}_t \boldsymbol{\Sigma}_{t|t-1} \tag{7}$$

where $\boldsymbol{K}_t = \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{x}_t' f_{t|t-1}$ is the Kalman gain. In a second step, H_t could be drawn by using the algorithm of Kim et al. (1998). Finally, one might draw $\boldsymbol{Q}_t$ as in Primiceri

[3] DMA averages over 2^m models, where m denotes the number of variables. For example, in the case m=10 this amounts to averaging over 1024 models.

[4] An index in the exponent denotes an information set.

[5] Note that, for simplicity, the model index (k) is removed temporarily.

(2005). Note that this algorithm depends on both, H_t and $\boldsymbol{Q}_t$. Running MCMC on K models, where each of these models may have many predictors, is computationally demanding. Therefore, we follow the suggestion in Raftery et al. (2010), which results in a proper approximation, such that the Kalman Filter only needs to be run K times. This approximation requires to set three hyperparameters λ, α and κ. The first two are called *forgetting factors* and the third a *decay factor*. Note that conditional on H_t, computation simplifies if equation (3) is replaced by

$$\boldsymbol{\Sigma}_{t|t-1} = \frac{1}{\lambda}\boldsymbol{\Sigma}_{t-1|t-1} \tag{8}$$

or, equivalently, $\boldsymbol{Q_t} = (\lambda^{-1} - 1)\boldsymbol{\Sigma}_{t-1|t-1}$, where $0 < \lambda \leq 1$.[6] Such approaches go back to Jazwinsky (1970). The term forgetting factor arises from the fact that observations i periods in the past obtain a weight of λ^i. E.g. if $\lambda = 0.99$ observations 12 months ago receive a weight of approximately 90%. To prevent the model from overfitting and to ensure that the coefficients evolve gradually over time, our analysis is limited to $\lambda \in \{0.99, 0.98\}$. Such values are common in the literature of forgetting factors, see e.g. Koop and Korobilis (2013), Di Filippo (2015) or Prüser (2019).[7] It is important to note that using (8) instead of (3) simplifies the MCMC algorithm since the need of estimating $\boldsymbol{Q_t}$ disappears. To fully circumvent the computational costs of MCMC-methods, a recursive estimator for H_t is needed. We follow the suggestion in Koop and Korobilis (2012) and use an Exponentially Weighted Moving Average (EWMA) such that

$$\hat{H}_{t|t-1} = \kappa \hat{H}_{t-1|t-2} + (1-\kappa)(y_t - \boldsymbol{x}_t\boldsymbol{\theta}_{t|t-1})^2, \tag{9}$$

where κ is the decay factor. EWMA estimators are frequently used in finance to model time-varying volatilities. Since monthly data are used in this application, we follow the suggestion in RiskMetrics (1996) and set $\kappa = 0.97$. The obvious advantage is that simulating H_t becomes obsolete and, in consequence, we do not need MCMC methods. Forecasting is then conducted by applying the predictive distribution

$$y_t|y^{t-1} \sim N(\boldsymbol{x}_t\boldsymbol{\theta}_{t|t-1}, \boldsymbol{x}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{x}_t' + H_t). \tag{10}$$

Finally, and in order to be able to average over K models, we need to calculate $P(M_t = k|y^{t-1})$. We again follow the suggestion in Raftery et al. (2010) and calculate time-

[6]In the special case of $\lambda = 1$, the marginal effects are time-invariant.

[7]Note that in the case of monthly data $\lambda = 0.98$ leads to a similar size of forgetting as using $\lambda = 0.95$ on quarterly data.

varying inclusion probabilities for each model as

$$\pi_{t|t-1,k} = \frac{\pi^{\alpha}_{t-1|t-1,k}}{\sum\limits_{l=1}^{K} \pi^{\alpha}_{t-1|t-1,l}} \tag{11}$$

where α is typically set to a value slightly less than one. The advantage of this specification is parsimony. That is, we do not need to specify a transition matrix which would again require MCMC methods. Finally, equation (11) is updated such that

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} p_k(y_t|y^{t-1})}{\sum\limits_{l=1}^{K} \pi_{t|t-1,l} p_l(y_t|y^{t-1})} \tag{12}$$

where $p_k(y_t|y^{t-1})$ is the predictive likelihood of model k. The predictive likelihood is defined as the predictive density evaluated at the actual observation y_t where the latter is given by equation (10). Forecasting is then done by weighting the predictive results of each model by using $\pi_{t|t-1,k}$. That is, DMA point predictions are given by

$$E(y_t|y^{t-1}) = \sum_{k=1}^{K} \pi_{t|t-1,k} \boldsymbol{x_t}^{(k)} \hat{\boldsymbol{\theta}}^{(k)}_{t|t-1}. \tag{13}$$

Dynamic Model Selection (DMS), instead, selects the single best model with highest $\pi_{t|t-1,k}$ at each point in time to make a prediction.
In order to provide a deeper understanding of the forgetting factor α, note that the weights assigned to each model k at time t are given by

$$\pi_{t|t-1,k} \propto \prod_{i=1}^{t-1} [p_k(y_{t-i}|y^{t-i-1})]^{\alpha^i}. \tag{14}$$

Thus, the weight attached to model k at time t depends on its forecasting performance in the recent past measured by the predictive likelihood $p_k(y_{t-i}|y^{t-i-1})$ and is controlled by the forgetting factor α. Therefore, the weight decays exponentially such that forecasts i periods in the past obtain a weight of α^i. E.g. if $\alpha = 0.99$, as in our benchmark scenario, forecasting performance 12 months in the past receives a weight of approximately 90%.

3. Design of the Forecasting Experiment

In order to evaluate the pseudo out-of-sample forecasting performance of the leading indicators, we follow the direct forecasting approach described in Marcellino et al. (2006). Let X_t denote the log level of the variable to be forecasted, in our example the log of the seasonally adjusted industrial production index. Furthermore, let y_t denote the stationary transformation of X_t after applying first differences. Since we are interested in annualized h-step ahead predictions, where $h \in \{1, 3, 6, 12\}$, the dependent variable in our forecasting exercise is y^h_{t+h} where

$$y^h_{t+h} = \frac{12}{h} X_{t+h} - X_t.$$

By including lags and exogenous regressors, the forecasting regression model becomes

$$y^h_{t+h} = c_t + \sum_{i=1}^{p} \rho_{t,i} y_{t+1-i} + \boldsymbol{x_t \theta_t} + \epsilon_{t+h},$$

where c_t denotes a constant term, $\rho_{t,i}$ the autoregressive coefficients, p the number of lags and $\boldsymbol{\theta_t}$ the effect of the m exogenous predictors $\boldsymbol{x_t}$. When applying DMA and DMS, a constant term and three lags enter each of the K models, regardless of h.[8] The lag length is motivated by the benchmark, a recursive AR(3) model, and has been determined by the AIC over the whole sample.[9] The prediction error of model j is given by $e^h_{j,t} = y_{t+h} - \hat{y}_{j,t+h}$.[10] To evaluate the predictive performance, we use two common forecast metrics, the root mean squared error (RMSE) and the mean absolute error (MAE) defined as

$$\text{RMSE}^h_j = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (e^h_{j,t})^2}$$

$$\text{MAE}^h_j = \frac{1}{T} \sum_{t=1}^{T} |e^h_{j,t}|.$$

[8]In an additional robustness check, the number of lags in each of the K models has been doubled. This did not improve the overall forecast accuracy. Furthermore, the variable inclusion probabilities stay the same. Results are available upon request.

[9]Using an AR process as the benchmark model is common in the literature on predicting industrial production in Germany, see e.g. Schumacher and Dreger (2005) or Wohlrabe and Robinzonov (2008).

[10]Note that $j \neq k$. Instead, j refers to, e.g., DMA for a certain combination of λ and α.

To allow for an immediate quantitative judgement of the actual forecasting model (j) against the benchmark model ($j = 0$), the relative forecasting performance is used. That is, we divide the RMSE (MAE) of the current forecasting model by the RMSE (MAE) of the benchmark model. If the resulting value is smaller than one, the current model outperforms the benchmark model. In addition, the predictive accuracy between the current forecasting model and the benchmark model is evaluated by applying the Diebold and Mariano (1995) test (DM-test). This test bases on the loss difference $d_{j,t} = g(e_{0,t}) - g(e_{j,t})$ where $g(.)$ can be any arbitrary function. Within our analysis, $g(.)$ is either the squared loss function or the absolute loss function depending on the forecasting metric at hand. The corresponding test-statistic is defined as

$$\text{DM} = \frac{\bar{d}_j}{\hat{V}(d_{j,t})}$$

where $\bar{d}_j$ and $\hat{V}(d_{j,t})$ are estimates of the mean and the long-run variance of $d_{j,t}$, respectively, and DM $\sim N(0,1)$. In certain circumstances, estimates of the long-run variance may become negative. Harvey et al. (2017) provide an overview about alternative estimators for the long-run variance in such circumstances. We follow one of their suggestions and make use of the Bartlett kernel. But Kiefer and Vogelsang (2005) show that the asymptotic distribution of such heteroskedasticity-autocorrelation (HAC) tests, based on nonparametric variance estimators, depends on the kernel and the bandwidth. Therefore, we make use of the Fixed-b asymptotic critical values suggested by them.[11]

4. Data

Within this out-of-sample forecasting exercise, we use monthly data ranging from June 1999 until October 2018. The first three years of data are used to initialize the recursive AR benchmark model. In addition, we lose three observations due to the lag length of all models. To forecast industrial production, 23 exogenous variables from three categories are considered.[12] The first category consists of twelve variables which typically exhibit *leading properties*. Within this study, we use three types of leading indicators. The first are survey based indicators and the second are composite indicators. While the former are immediately derived from survey data, the latter is an aggregated index comprising

[11]The interested reader is referred to Kiefer and Vogelsang (2005) for further details.
[12]The exact choice of the variables is a distillate of the relevant literature.

individual indicators.[13] Four out of six indicators used in this study, namely the ifo business climate, the ZEW indicator, the consumer confidence indicator and the business confidence indicator, are survey based while the two remaining indicators, namely the composite leading indicator from the OECD and the euro-coin indicator, belong to the class of composite indicators. The third type includes various new orders time series for industry, intermediate goods, investment, consumption, domestic and foreign.
The second category consists of five *macroeconomic* time series, namely the unemployment rate, EONIA, harmonized consumer price index, labor force and employees with social insurance. Six *financial* variables constitute the third category, namely M3, a commodity price index, the DAX, an interest rate spread (10 year government bond - EONIA), the effective exchange rate and a financial stress indicator. Table 1 provides an overview about the data sources and the transformations applied to each time series. Since the survey based and composite indicators are approximately stationary in levels, one might be tempted to include them in levels. But Wohlrabe and Wollmershäuser (2017) show for both, the ifo and the ZEW indicator, that the correlation between the rates of change of GDP and first differences of these indicators is higher. Therefore, we follow this suggestion.[14]

We provide a comprehensive overview about the forecasting properties of all three variable categories by estimating five different models. Each models' ingredients are provided in Table 1. Model 1 (M1) consists of 18 variables from all three categories. That is, we consider the whole range of survey and data based leading indicators, several new orders time series, three macroeconomic indicators and all six financial indicators.[15] The remaining four models are, with respect to the included variable categories, in principle subsets of M1. Model 2 (M2) considers only survey based and composite indicators.[16] Model 3 (M3) extends M2 by the new orders indicators. Model 4 (M4) considers exclusively macroeconomic variables and Model 5 (M5) solely financial indicators. One might question the need for M2 until M5 since our econometric framework considers variable uncertainty such that less important predictors receive little weight. However, only estimating M1 will not provide information regarding the quantitative predictive power of

[13]A comprehensive overview about the construction of composite indicators is given by Nardo et al. (2005).

[14]In a later robustness check the leading indicators enter the model in levels.

[15]We did not consider estimating a model with all predictors, because computation time of this model increases exponentially. Furthermore, the correlation coefficients between all new orders series are very high.

[16]Both composite indicators, the CLI and the EUC, are omitted in a later robustness check.

Table 1: Variable and Model Overview

Category	Variable	M1	M2	M3	M4	M5	Abbr.	Source	T-Code
Endogenous	Industrial Production	x	x	x	x	x	IP	Buba	4
Leading Indicator	Ifo Business Climate	x	x	x			IFO	CESifo	3
	ZEW	x	x	x			ZEW	ZEW	3
	Consumer Confidence	x	x	x			CCI	EC	3
	Business Confidence	x	x	x			BCI	EC	3
	Composite Leading	x	x	x			CLI	OECD	3
	Euro-coin	x	x	x			ECI	CEPR	3
	New Orders Industry	x		x			NOIN	Buba	4
	New Orders Intermediate			x			NOINP	Buba	4
	New Orders Investment			x			NOINV	Buba	4
	New Orders Consumption	x		x			NOCO	Buba	4
	New Orders Domestic			x			NODOM	Buba	4
	New Orders Foreign	x		x			NOFO	Buba	4
Macro	Unemployment Rate	x			x		UN	BFA	3
	EONIA	x			x		EON	EZB	1
	Consumer Prices	x			x		HICP	Buba	4
	Labor Force				x		LF	Buba	4
	Social insured Employees				x		SILF	Buba	4
Finance	M3	x				x	M3	Buba	4
	Commodity Price Index	x				x	COM	HWWI	4
	Dax	x				x	DAX	YF	4
	10 Year - EONIA	x				x	SPREAD	OC	1
	Effective Exchange Rate	x				x	FX	EZB	1
	Financial Stress Indicator	x				x	FS	EZB	2

Notes: T-Code: 1 - Level; 2 - Log Level; 3 - First Difference; 4 - Log Difference. BFA - Bundesagentur für Arbeit, YF - Yahoo Finance, OC - Own Calculation.

Table 2: Correlation coefficients between leading indicators (in differences)

	ifo	ZEW	CCI	BCI	CLI	EUC
ifo	1.00					
ZEW	0.14	1.00				
CCI	0.46	0.14	1.00			
BCI	0.73	0.06	0.39	1.00		
CLI	0.71	0.16	0.44	0.75	1.00	
EUC	0.50	0.22	0.33	0.47	0.33	1.00

all three categories. Estimating all five models allows for an immediate comparison of the quantitative predictive power across all three categories.

Before we present our empirical results, we briefly sketch the construction of the survey based and composite indicators. The ifo business climate bases on 9.000 monthly reports of the manufacturing, service, building and retail branch. It is an average of the questions about the current business situation and about the expectation about the next six months. The ZEW indicator uses a survey of 300 experts from banks, insurance companies and finance departments. More explicitly, the index bases on their expectations about the business cycle in six months. In order to capture the expectations of households, the consumer confidence indicator utilises information from a survey of 2.000 persons. Those individuals are asked about their expected financial situation, the general economic situation, prices, unemployment as well as about their consumption and savings behaviour. The business confidence indicator, calculated by the European Commission, uses the same questionnaire as the ifo business climate. However, it utilises information from different questions. This indicator combines information about the stock of finished goods, new orders as well as about domestic production within the next three months. Compared to the just mentioned survey based indicators, the composite leading indicator is a weighted average of the ifo business climate indicator, orders inflow, export order books, new orders in manufacturing, finished goods stock, spread of interest rates, service demand evolution and the consumer confidence indicator. Therefore, it draws on a wider information set compared to the survey based indicators and provides a good approximation of the business cycle. Finally, the euro-coin indicator bases on generalized principal components and condenses information from 145 monthly time series, from various countries, to a single indicator of economic activity for the Euro area. Therefore, our analysis does not only employ information from Germany, but also from the Euro area. Utilizing such information is especially important for foreign trade dependent countries like Germany. Table 2 provides an overview about the correlation between all six leading indicators. While the pairwise correlations between the business confidence indicator, the composite leading indicator, the ifo index and the euro-coin are relatively high (around 0.7, except between BCI and EUC) the ZEW indicator is, surprisingly, only weakly correlated with all other indicators.

5. Empirical results

This section presents the empirical results. We begin with the quantitative forecasting results of M1 until M5. Next, the posterior inclusion probabilities of M1 are shown, followed by a discussion about the average model size considered at each point in time. Finally, we investigate the evolution of the cumulated absolute forecast error to reveal periods with poor forecast performance.

The relative forecast performance of all models is summarized in Table 3. The core results base on the forgetting factors $\lambda = 0.99$ and $\alpha = 0.99$ in order to favor gradual changes of the model and its parameters.[17] The relative forecast performance is provided based on both, the RMSE- and the MAE-metric. As a first step, we compare the predictive power of the benchmark model with an autoregressive time-varying parameter model.[18] The results show that allowing for time-variation increases the predictive power slightly. Focussing on the quantitative forecasting results of M1 until M3 in terms of the relative forecasting performance leads to the following conclusions.[19] Both, DMA and DMS beat the benchmark model at all four horizons independent from the model's ingredients since the relative forecast performance is below one.[20] The improvements range between -3% and -33%. In most instances, the DM-test reveals significant outperformance of DMA and DMS against the benchmark model. The forecast performance of these three models, conditional on h, is very similar. The only exception arises at $h = 12$. The relative forecasting performance of M1 collapses to 0.94 while the relative forecasting performances of M2 and M3 is approximately 0.77 when considering the RMSE.

[17] This assumption is relaxed in a later sensitivity analysis.

[18] The forgetting factor of the autoregressive time-varying parameter has been set to $\lambda = 0.99$.

[19] Note that all three models contain variables from the class of leading indicator.

[20] There are very few instances in which a contrary result arises and these are limited to M1 when using the MAE-metric with a forecast horizon of twelve months. Note that this result depends on the specification of α. In a later robustness check we consider a higher degree of model change.

Table 3: Relative Forecast Performance: $\lambda = 0.99$, $\alpha = 0.99$

Metric	h	AR-TVP (0.99)	M1		M2		M3		M4		M5	
			DMA	DMS	DMA	DMS	DMA	DMS	DMA	DMS	DMA	DMS
RMSE	1	0.99**	0.91**	0.92**	0.90**	0.90**	0.91**	0.92*	0.99	1.00	0.97**	0.98
	3	0.93*	0.71*	0.72*	0.74**	0.75*	0.73*	0.75*	0.95	0.95	0.90*	0.90*
	6	0.96	0.69**	0.69**	0.67**	0.68**	0.67**	0.68**	0.97	0.97	0.87**	0.87*
	12	1.01	0.94	0.95	0.77***	0.77***	0.77***	0.77***	1.07	1.07	1.02	1.00
MAE	1	0.99*	0.97	0.97	0.95*	0.95	0.96	0.97	0.99	0.99	1.00	1.02
	3	0.95**	0.84**	0.85*	0.81**	0.83**	0.82**	0.83**	0.97	0.97	0.97	0.99
	6	0.93**	0.76**	0.77**	0.70***	0.71***	0.70***	0.71***	1.01	1.02	0.94	0.96
	12	0.99	1.02	1.06	0.73***	0.73***	0.73***	0.73***	1.19	1.19	1.07	1.06

Notes: Metric determines one of the two forecast metrics, the Root Mean Squared Error (RMSE) or the Mean Absolute Error (MAE); $h \in \{1, 3, 6, 12\}$ corresponds to the forecast horizon; *** (**/*) denotes significant outperformance of the respective model against an AR(3) at the 1 (5/10)% level using the one-sided Diebold-Mariano-Test. The critical values have been calculated by using the Fixed-b asymptotic derived in Kiefer and Vogelsang (2005). AR-TVP (0.99) is a simple time-varying AR model where $\lambda = 0.99$.

Next, we focus our attention on M4 and M5 considering solely macroeconomic or financial variables, respectively. When applying the RMSE, the relative forecasting performance of M4 is below one for $h \in \{1, 3, 6\}$ while for the MAE, it is only below one if $h \in \{1, 3\}$. Therefore, macroeconomic variables are relatively weak predictors, especially at a longer forecasting horizon, e.g. at $h = 12$. The null hypothesis of equal forecast accuracy is not rejected in almost all instances. The pattern changes slightly when considering the relative forecasting performance of M5. In most cases the relative forecast performance is below one. But more important, the relative forecast performance is better than in M4. Additionally, the DM-test confirms the outperformance at several horizons.
Summing up, those models considering leading indicators, M1 until M3, beat the benchmark model at all horizons. The relative forecast performance of these three models, conditional on h, is very similar. Comparing the relative forecast performance of M2 and M3, where M2 is basically a constrained version of M3, yields that the additional explanatory power of the new orders series is limited. M4 and M5, the models uniquely considering either macroeconomic or financial variables, forecast slightly better than the benchmark model. But they exhibit a weaker predictive performance compared to M1 until M3. Furthermore, macroeconomic variables forecast worse compared to financial variables. The finding that the relative forecast performance of M1, M2 and M3 decreases with the forecast horizon h is in line with Drechsel and Scheufele (2012).[21]

So far, we have only focused on the quantitative forecasting performance of each model against the benchmark model and additionally on the forecasting performance of these models with partly non-overlapping categorial ingredients. However, this perspective ignores one important component, namely the weight attached to each predictor. For example, a single predictor of a certain category might have a lot of predictive power while the remaining variables are mostly useless. One attractive feature of our econometric framework is that it allows to compute weights for each single predictor. Such weights are typically called inclusion probabilities. These probabilities are the sum of the weights of all models which include a particular predictor. This feature of the model allows us to detect the variable with the highest weight. Furthermore, the weight of each predictor might change along two dimensions, over time and with respect to the forecast horizon. DMA is able to detect such circumstances. For example, some leading

[21] This finding is due to the fact that leading indicators deploy their predictive power at longer forecast horizons.

indicators might forecast well if h is small and some if h is large. Obviously, these indicators should be attached with higher weight when h is small or large. We will focus on the time-varying variable probabilities derived from M1, since the weights are of special interest when considering a model with many competing predictors.[22]

Figure 1 provides an overview about the inclusion probabilities and consists of twelve sub-plots. The rows correspond to the forecast horizon h while the three columns display the time-varying variable probabilities for the 18 candidate predictors. We begin with the first row of Figure 1 depicting the time-varying variable probabilities when forecasting one period ahead ($h = 1$). The first sub-plot shows the inclusion probabilities of the survey and composite indicators. The weights of most leading indicators stay below 50%. The only exceptions are BCI and CLI. The former has an inclusion probability of more than 60% over the whole period under investigation while the latter has the second highest weight. This pattern changes after 2014. The inclusion probability of CLI exceeds that of BCI. The second sub-plot shows the inclusion probabilities of NOIN, NOCO, NOFO, UN, EON and HICP. The weights of all variables are quite similar. NOFO and HICP have, taking an average over time, the highest inclusion probability. But none of these series exceeds the probability of 50%. Finally, we turn our attention to the financial variables. Their inclusion probabilities are again quite similar. All of them rarely exceed the 50% threshold. In summary, for making forecasts one month ahead ($h = 1$) the leading indicators BCI and CLI are most useful.

When predicting industrial production three months ahead (row 2 of Figure 1), the pattern starts to change. The first sub-plot reveals that the weight attached to several leading indicators varies strongly over time. The weight attached to CLI is now higher than the weight attached to BCI along almost the entire sample period under investigation for predicting industrial production in Germany. Furthermore, the weight attached to CLI is 100% since the onset of the Great Recession. In addition, the weight attached to ZEW and EUC increases at the beginning of the Great Recession and stays relatively high over the remaining sample. The inclusion probabilities in the second sub-plot, those of new orders series as well as of the macroeconomic variables, are on average smaller compared to $h = 1$. Finally, the third sub-plot reveals that both, FS and COM, receive the highest weight from 2008 onwards while the weight of the remaining variables stays below 40% over the entire sample.

[22]We have also calculated the time-varying variable probabilities for M2 until M5. The results are available upon request.

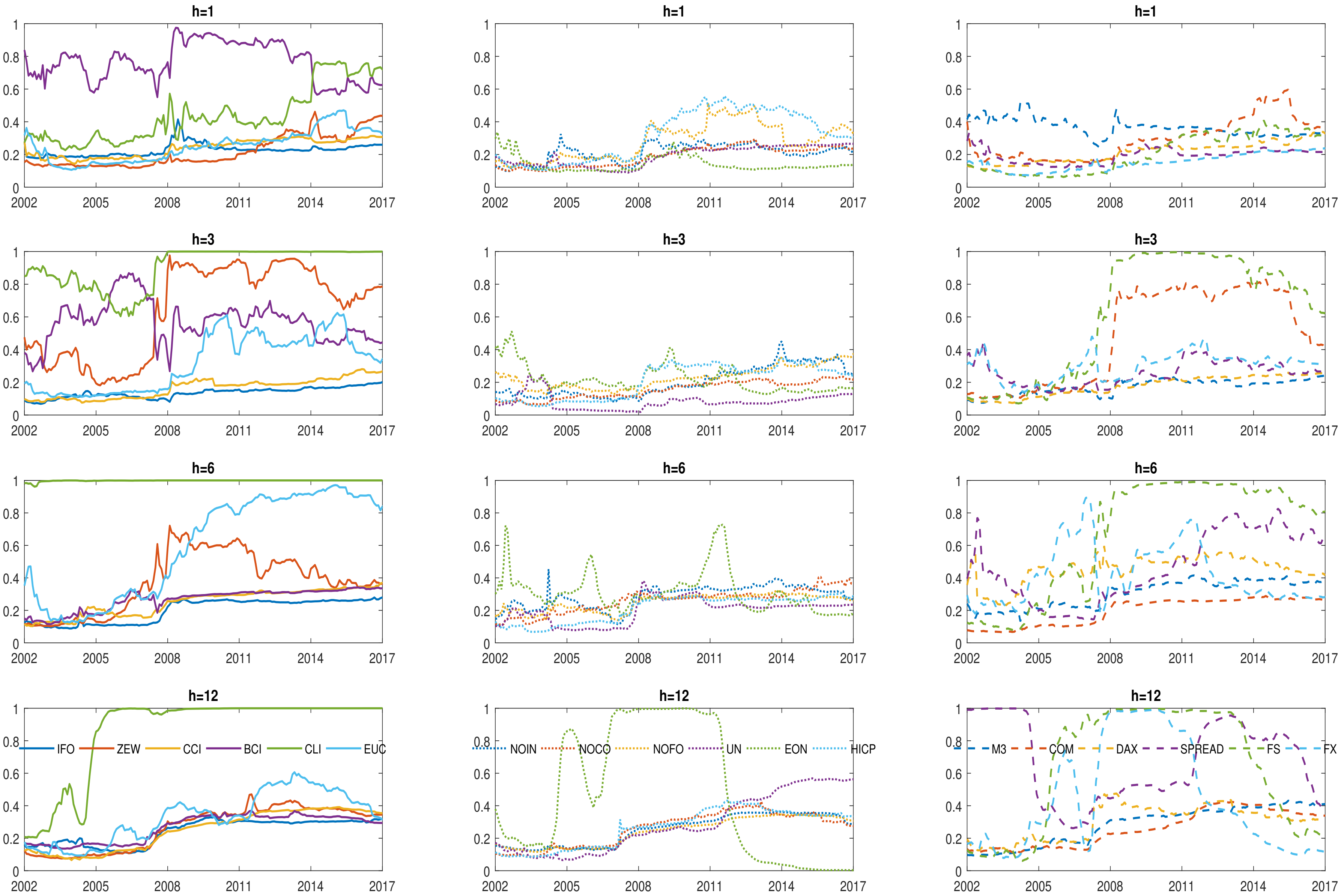

Figure 1: Posterior Inclusion Probabilities of M1, $\lambda = 0.99$, $\alpha = 0.99$.

The picture again changes slightly when moving to the third row, which provides time-varying inclusion probabilities when forecasting six periods ahead ($h = 6$). Interestingly, DMA puts a weight of almost one on CLI over almost the entire sample. Additionally, the weight attached to EUC increases with the onset of the Great Recession, while the weight of the ZEW indicator decreases compared to the three month horizon. This increase might be caused by the substantial economic spillovers during the sovereign debt crisis. The remaining indicators receive a weight smaller than approximately 30%. The inclusion probabilities in the second sub-plot are almost equal to those of the three period horizon. The only exception is EON which spikes at three points in time. Lastly, we focus on the inclusion probabilities of the financial variables. Compared to the three months horizon, FS is of similar importance while the weight on COM has decreased strongly. Two further variables exceed the threshold of 50% at some points in time, namely the SPREAD and FX.

Lastly, we investigate the time-varying inclusion probabilities for a forecasting horizon of twelve months (fourth row). The first sub-plot reveals that DMA still puts the largest weight on CLI while the weights on the remaining indicators are quite low. Within the variable set depicted in the second sub-plot, only EON receives a relatively high weight. Finally, the third sub-plot shows that SPREAD, FS and FX experience periods with relatively high weights.

Summing up, there are considerable changes in the inclusion probabilities along both dimensions, time and forecast horizon. At the one month forecast horizon, BCI and CLI are the predictors with the largest weight. Especially BCI seems to be an important predictor in the short run. When forecasting three months ahead, one should consider CLI, ZEW, BCI, EUC, FS and COM. At the six months horizon, the weights attached to BCI, ZEW, COM decrease while the weights on EUC and SPREAD increase. Finally, on the twelve months horizon, only CLI, EON, SPREAD, FS and FX are equipped with a relatively high weight. The high weight attached to CLI when forecasting three, six and twelve months ahead fits roughly to the lead length targeted by Gyomai and Guidetti (2012) ranging between six to nine months. In the case of the EON, the high weight attached at $h = 12$ might be due to the lag with which monetary policy affects the economy. The sudden decline of the weight since 2013 may be due to the fact that the EON has already reached a very low level. Overall, the results confirm the need of a dynamic estimator where the set of candidate predictors receives time-varying weights.

In the following, we provide evidence regarding the need of an econometric framework

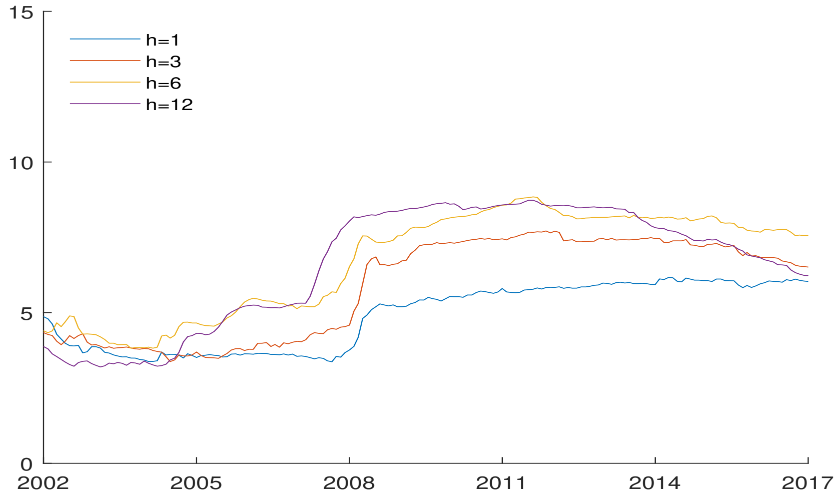

Figure 2: Predicted Model Size of M1 over time and in relation to the forecasting horizon, $\lambda = 0.99$, $\alpha = 0.99$.

that can deal with the issue of model size in a dynamic fashion. Let $Size_{k,t}$ denote the number of predictors in model k at time t.[23] Then

$$E(Size_t) = \sum_{k=1}^{K} \pi_{t|t-1,k} Size_{k,t}$$

can be interpreted as the average number of predictors used in DMA at time t. Figure 2 illustrates this measure for M1 conditional on h. It is remarkable that the full set of predictors is never considered. Instead, DMA favours more parsimonious models. At the beginning of the sample, the average model size is slightly below five. Moving along the time line, the average model size starts to increase with the onset of the Great Recession. Therefore, during times of economic turmoil, DMA favours additional regressors or, put differently, more complex models. Not unsurprisingly, the average number of predictors of models with large h seem to experience this increase somewhat earlier. Immediately after the Great Recession the average number of predictors stays the same and starts to decline for $h \in \{3, 6, 12\}$ at approximately 2012. Furthermore, Figure 2 clearly highlights that multi-predictor models are superior compared to single-predictor models, which are usually considered in many empirical studies.

Next, we focus on the predictive performance of all five models over time by investigating the cumulated absolute forecast error. This allows us to detect structural breaks in the

[23]Note that $Size_{k,t}$ does not contain the intercept as well as the three lags included in each model.

forecasting performance. Figure 3 provides the corresponding overview. The figure consists of 20 sub-plots. The rows correspond to the forecast horizon $h \in \{1, 3, 6, 12\}$ and the columns to M1 until M5. Within each sub-panel, the cumulated absolute forecast errors are depicted. The blue line corresponds to the benchmark model, the red line to the AR-TVP(0.99) model and the yellow and purple line to DMA and DMS, respectively. A very similar pattern is observed in many sub-plots. That is, a steady increase of the cumulated absolute forecast error until the beginning of the Great Recession followed by a sudden jump and finally a steady increase in the aftermath of the Great Recession. Therefore, the largest forecasting errors are made during times of economic turmoil.

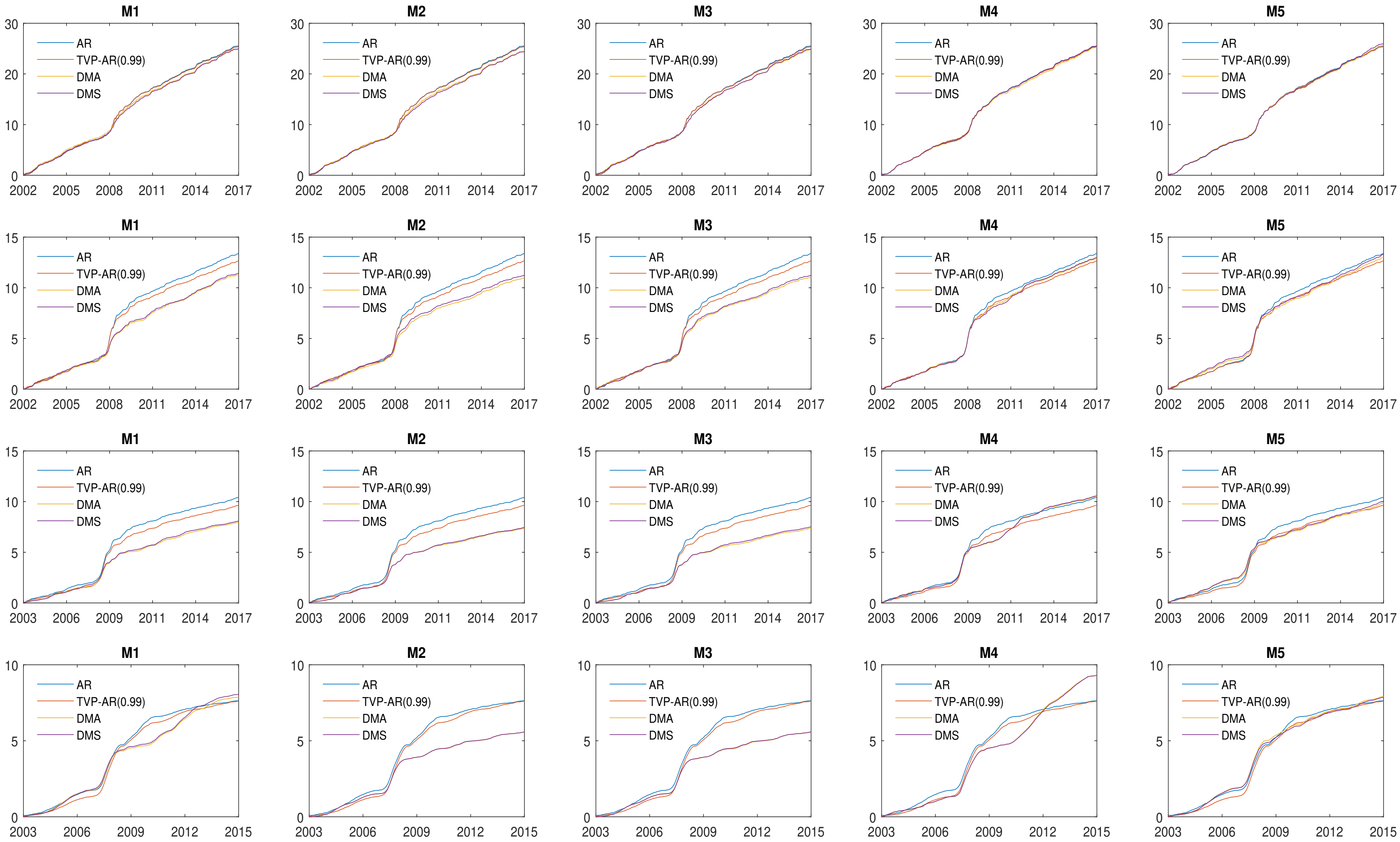

Figure 3: Cumulated absolute forecast error for each model. The four rows correspond to $h \in \{1, 3, 6, 12\}$.

6. Robustness

We investigate the robustness of our results in two ways. In a first step, we focus on the sensitivity of the results with respect to the two forgetting factors λ and α. In a second step, we focus on the forecasting properties of M2, the leading indicators. Since we are primarily interested in their forecasting abilities, we focus on M2 and consider the following scenarios. We evaluate the forecasting performance of those indicators in level (Abbr. LiL in Table A.1). Furthermore, we solely consider survey based indicators (Abbr. SV in Table A.1). That is, we do not consider the CLI and EUC when forecasting industrial production in Germany. This is due to the fact that composite indicators are usually subject to revisions and may be published with a delay due to data availability.

We start with the sensitivity analysis where $\lambda = 0.98$ and $\alpha = 0.98$. Table A.2 provides an overview. In principle, there are no major differences in the quantitative results. Figure A.2 depicts the posterior inclusion probabilities. The major conclusions drawn stay the same even though the inclusion probabilities are more erratic compared to the specification above. Furthermore, the average number of predictors used at time t, as depicted in Figure A.1 in general tells the same story. The only differences are that the average model size is somewhat larger along the entire sample and that it decreases more quickly after the Great Recession.
Finally, we focus attention on the robustness checks for M2 to be found in Table A.1. Using all leading indicators in levels (see column LiL in Table A.1) leads to worse predictions compared to first differences. Finally, when removing the two data based indicators, the forecast quality worsens quite strongly (see column SV in Table A.1). However, this does not come as a surprise since our empirical investigation has shown that the weight attached to CLI is very large independent from the forecast horizon.

7. Conclusions

This study uses Dynamic Model Averaging to investigate the predictive power of leading indicators within a pseudo out-of-sample forecasting exercise. Furthermore, we challenge them with macroeconomic and financial variables when forecasting industrial production in Germany. DMA has three important and attractive econometric properties. That is, it allows the model coefficients to change over time, it deals with model selection appropriately and, finally, allows the entire forecasting model to change over time. Our analysis shows that the weight attached to all three variable categories is time-varying and that variables from the category of leading indicators receive the largest weight independent from the forecast horizon. Leading indicators show the best quantitative forecasting properties for industrial production while both, macroeconomic and financial variables are less informative. We find that the model size changes considerably over time, especially since the onset of the Great Recession. That is, during this period of massive economic turmoil, larger forecasting models receive a higher weight. Furthermore, there seems to be a positive association between the model size and the forecast horizon. Finally, the cumulated absolute forecast error points towards similar forecast performance before and after the Great Recession.

Many time series are subject to revisions. This implies that information available earlier may differ from later-available information for a certain point in time. Researchers promoting the use of real-time data in forecasting comparisons usually argue that models should be evaluated with data available at the point in time when the forecasts are made. Indeed, when comparing the performance of econometric models with forecasts from professionals this issue is uncontroversial. However, the consequences for model/variable selection and evaluation are less clear. This issue has been investigated by Heinisch and Scheufele (2018). They find that the relative forecasting performance does not depend on the use of real-time or final data.[24] Furthermore, using real-time data is not without costs. It increases the number of time series to be used substantially and therefore causes enormous computational costs, especially when DMA is used. Therefore, more efficient methods are needed. One solution could be the use of dynamic Occam's window as suggested by Onorante and Raftery (2016). The idea behind this extension is to decrease the computational burden by focusing on models whose weights exceed a certain threshold. However, we leave this issue for future research.

[24] Bernanke and Boivin (2003) and Schumacher and Breitung (2008) also have shown that there is no major difference between forecasts from final and real-time data.

References

Barnett, A., Mumtaz, H., and Theodoridis, K. (2014). Forecasting UK GDP growth and Inflation under Structural Change. A Comparison of Models with time-varying Parameters. *International Journal of Forecasting*, 30(1):129–143.

Bernanke, B. S. and Boivin, J. (2003). Monetary Policy in a data-rich Environment. *Journal of Monetary Economics*, 50(3):525–546.

Chan, J. C. C., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2012). Time-varying Dimension Models. *Journal of Business and Economic Statistics*, 30(3):358–367.

Cogley, T., Morozov, S., and Sargent, T. J. (2005). Bayesian fan charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an evolving Monetary System. *Journal of Economic Dynamics and Control*, 29(11):1893–1925.

Di Filippo, G. (2015). Dynamic Model Averaging and CPI Inflation Forecasts: A Comparison between the Euro Area and the United States. *Journal of Forecasting*, 34(8):619–648.

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

Drechsel, K. and Scheufele, R. (2012). The Performance of short-term Forecasts of the German Economy before and during the 2008/2009 Recession. *International Journal of Forecasting*, 28(2):428–445.

Ferrara, L., Marcellino, M., and Mogliani, M. (2015). Macroeconomic forecasting during the great recession: The return of non-linearity. *International Journal of Forecasting*, 31(3):664–679.

Groen, J., Paap, R., and Ravazzolo, F. (2009). Real-Time Inflation Forecasting in a Changing World. Technical Report May, Federal Reserve Bank of New York.

Gyomai, G. and Guidetti, E. (2012). OECD System of Composite Leading Indicators. Technical report available at: https://www.oecd.org/sdd/leading-indicators/41629509.pdf.

Harvey, D., Leybourne, S., and Whitehouse, E. (2017). Forecasting evaluation tests and negative long-run variance estimates in small samples. *International Journal of Forecasting*, 33(4):833–847.

Heinisch, K. and Scheufele, R. (2018). Should Forecasters use Real-Time Data to evaluate Leading Indicator Models for GDP Prediction? German Evidence. *German Economic Review*.

Hendry, D. F. and Clements, M. P. (2004). Pooling of Forecasts. *The Econometrics Journal*, 7(1):1–31.

Jazwinsky, A. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Kiefer, N. and Vogelsang, T. (2005). A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests. *Econometric Theory*, 21:1130–1164.

Kim, S., Shepard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65:361–393.

Koop, G. and Korobilis, D. (2012). Forecasting Inflation Using Dynamic Model Averaging. *International Economic Review*, 53(3):867–886.

Koop, G. and Korobilis, D. (2013). Large time-varying Parameter VARs. *Journal of Econometrics*, 177(2):185–198.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A Comparison of direct and iterated multistep AR methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics*, 135(1-2):499–526.

Nardo, M., Saisana, M., Saltelli, A., and Tarantola, S. (2005). Tools for Composite Indicators Building. *European Commission, Applied Statistics Group*.

Onorante, L. and Raftery, A. (2016). Dynamic Model Averaging in large Model Spaces using dynamic Occam's window. *European Economic Review*, 81:2–14.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions. *Review of Economic Studies*, 72(3):821–852.

Prüser, J. (2019). Adaptive Learning from Model Space. *Journal of Forecasting*, 38(1):29–38.

Raftery, A. E., Karny, M., and Ettler, P. (2010). Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics*, 52(1):52–66.

RiskMetrics (1996). *RiskMetrics - Technical Document.* J.P. Morgan, New York, 4. edition.

Schumacher, C. and Breitung, J. (2008). Real-time Forecasting of German GDP based on a Large Factor Model with monthly and quarterly Data. *International Journal of Forecasting*, 24(3):386–398.

Schumacher, C. and Dreger, C. (2005). Out-of-sample Performance of Leading Indicators for the German Business Cycle. *Journal of Business Cycle Measurement and Analysis*, 2(1):71–87.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.

Timmermann, A. (2006). Chapter 4: Forecast Combinations. In Elliott, G., Granger, C. W., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 135–196. Elsevier B.V.

Ulbricht, D., Kholodilin, K. A., and Thomas, T. (2017). Do Media Data Help to Predict German Industrial Production? *Journal of Forecasting*, 36:483–496.

Wohlrabe, K. and Robinzonov, N. (2008). Freedom of Choice in Macroeconomic Forecasting: An Illustration with German Industrial Production and Linear Models. *Ifo Working Paper*, (57).

Wohlrabe, K. and Wollmershäuser, T. (2017). Über die richtige Interpretation des ifo Geschäftsklimas als konjunktureller Frühindikator. *ifo Schnelldienst*, 70(15):42–46.

A. Appendix

Table A.1: Forecast Errors: $\lambda = 0.99$, $\alpha = 0.99$

Robustness		LiL		SV	
Metric	h	M2		M2	
		DMA	DMS	DMA	DMS
RMSE	1	0.98	1.01	0.90**	0.90**
	3	0.93	0.93	0.78**	0.79**
	6	0.93	0.94	0.82***	0.82***
	12	0.99	0.98	0.90***	0.90***
MAE	1	0.97	1.01	0.95	0.95
	3	0.93	0.94	0.87***	0.87***
	6	0.97	0.98	0.80***	0.80***
	12	0.96	0.97	0.85***	0.87***

Notes: Metric determines one of the two forecast metrics: Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE); $h \in \{1, 3, 6, 12\}$ corresponds to the forecast horizon; *** (**/*) denotes significant outperformance of the model against an AR(3) at the 1 (5/10)% level using the one-sided Diebold-Mariano-Test. The critical values have been calculated by using the Fixed-b asymptotic derived in Kiefer and Vogelsang (2005). LiL - Leading Indicator in Level; SV - considers only survey based indicators.

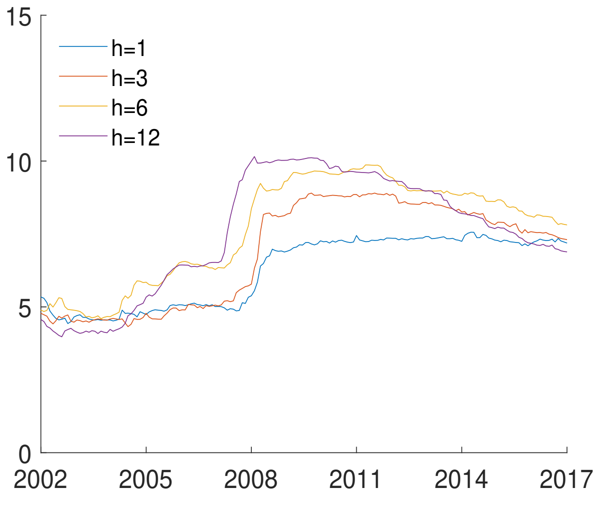

Figure A.1: Predicted Model Size of M1 over time and in relation to the forecast horizon, $\lambda = 0.98$, $\alpha = 0.98$.

Table A.2: Forecast Errors: $\lambda = 0.98$, $\alpha = 0.98$

Metric	h	AR-TVP (0.99)	M1		M2		M3		M4		M5	
			DMA	DMS	DMA	DMS	DMA	DMS	DMA	DMS	DMA	DMS
RMSE	1	0.99**	0.91**	0.91**	0.91**	0.92**	0.91**	0.94	0.98	0.99*	0.95*	0.96
	3	0.93*	0.72*	0.72*	0.76*	0.76*	0.75*	0.77*	0.96	0.98	0.89*	0.88
	6	0.96	0.74*	0.75*	0.72**	0.71**	0.72**	0.72*	0.96	0.96	0.92*	0.92*
	12	1.01	0.93	0.96	0.79***	0.79***	0.80***	0.79***	1.05	1.06	1.07	1.07
MAE	1	0.99*	0.97	0.98	0.96	0.97	0.97	1.00	0.99	1.00	1.00	1.01
	3	0.95**	0.85*	0.87	0.84**	0.85**	0.83**	0.85**	0.98	1.01	0.97	0.98
	6	0.93**	0.81**	0.86*	0.74**	0.75**	0.74**	0.76**	1.00	1.00	0.97	0.98
	12	0.99	0.99	1.07	0.74***	0.74***	0.74***	0.74***	1.11	1.12	1.09	1.10

Notes: Metric determines one of the two forecast metrics: Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE); $h \in \{1, 3, 6, 12\}$ corresponds to the forecast horizon; *** (**/*) denotes significant out performance of the Model against an AR(3) at the 1 (5/10)% level using the one-sided Diebold-Mariano-Test. The critical values have been calculated by using the Fixed-b asymptotic derived in Kiefer and Vogelsang (2005).

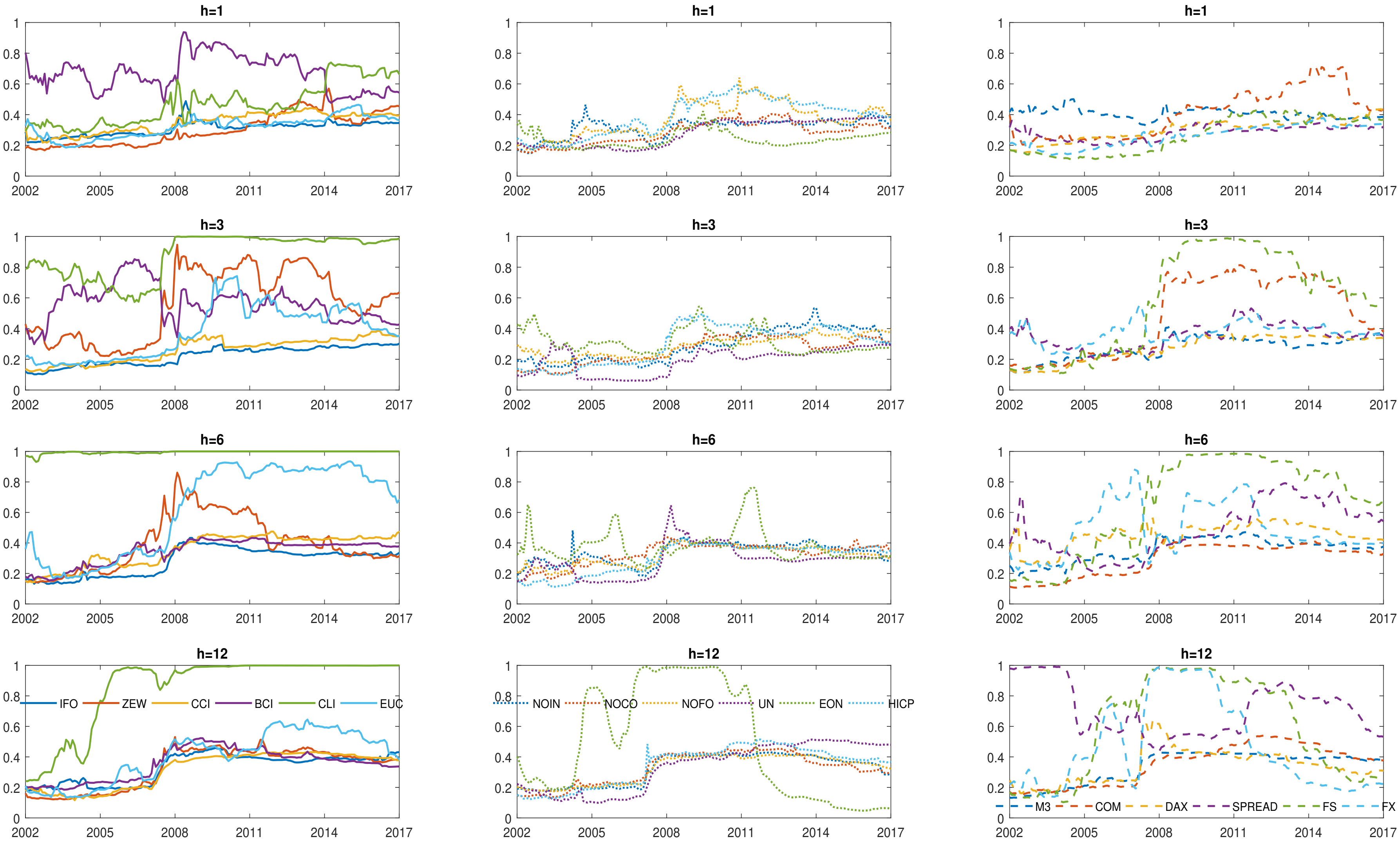

Figure A.2: Posterior Inclusion Probabilities of Model 1, $\lambda = 0.98$, $\alpha = 0.98$.