

Heinisch, Katja; Scheufele, Rolf

Working Paper

Should forecasters use real-time data to evaluate leading indicator models for GDP prediction? German evidence

IWH Discussion Papers, No. 5/2017

Provided in Cooperation with:

Halle Institute for Economic Research (IWH) – Member of the Leibniz Association

Suggested Citation: Heinisch, Katja; Scheufele, Rolf (2017) : Should forecasters use real-time data to evaluate leading indicator models for GDP prediction? German evidence, IWH Discussion Papers, No. 5/2017, Leibniz-Institut für Wirtschaftsforschung Halle (IWH), Halle (Saale), <https://nbn-resolving.de/urn:nbn:de:gbv:3:2-67732>

This Version is available at:

<https://hdl.handle.net/10419/149886>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Halle Institute for Economic Research
Member of the Leibniz Association

Discussion Papers

No. 5

February 2017



 Should Forecasters Use Real-time Data to Evaluate
Leading Indicator Models for GDP Prediction? German Evidence

Katja Heinisch, Rolf Scheufele

Authors

Katja Heinisch

Halle Institute for Economic Research (IWH) –
Member of the Leibniz Association
Department of Macroeconomics
E-mail: katja.heinisch@iwh-halle.de
Tel +49 345 7753 836

Rolf Scheufele

Swiss National Bank (SNB), and Halle
Institute for Economic Research (IWH) –
Member of the Leibniz Association
Department of Macroeconomics
E-mail: rolf.scheufele@snb.ch
Tel +41 58 631 3744

The responsibility for discussion papers lies solely with the individual authors. The views expressed herein do not necessarily represent those of the IWH. The papers represent preliminary work and are circulated to encourage discussion with the authors. Citation of the discussion papers should account for their provisional character; a revised version may be available directly from the authors.

Comments and suggestions on the methods and results presented are welcome.

IWH Discussion Papers are indexed in RePEc-EconPapers and in ECONIS.

Editor

Halle Institute for Economic Research (IWH) –
Member of the Leibniz Association

Address: Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

Should Forecasters Use Real-time Data to Evaluate Leading Indicator Models for GDP Prediction? German Evidence*

Abstract

In this paper we investigate whether differences exist among forecasts using real-time or latest-available data to predict gross domestic product (GDP). We employ mixed-frequency models and real-time data to reassess the role of survey data relative to industrial production and orders in Germany. Although we find evidence that forecast characteristics based on real-time and final data releases differ, we also observe minimal impacts on the relative forecasting performance of indicator models. However, when obtaining the optimal combination of soft and hard data, the use of final release data may understate the role of survey information.

Keywords: mixed-frequency VAR, real-time data, nowcasting, forecasting

JEL Classification: C53, C55, E37

* We would like to thank Alain Galli as well as the participants of the Annual International Symposium on Forecasting 2015, CFE 2015 and German Economic Association 2016 for helpful comments and suggestions.

1 Motivation

Macroeconomic forecasting faces many serious challenges. A major challenge in forecasting real activity is that both target and indicator variables are subject to data revisions, which implies that early available information (e.g., first GDP releases or early available predictors for GDP) may differ from later-available information for a specific point in time. This variation presents important implications for evaluating forecasts from different models because forecasts typically depend strongly on past observations via parameter estimates and starting levels.

In practice, macroeconomic forecasters often ignore the fact that data used for conducting forecasts may be revised from time to time. It is common practice to use final data — the latest-available data vintage of a given variable — to specify and evaluate a particular model. In this case, forecast comparisons are conducted by so-called pseudo out-of-sample forecast exercises that use the latest-available data recursively for model selection, model averaging, and model evaluation (e.g., Stock and Watson, 2003; Banerjee, Marcellino, and Masten, 2005). Thus, researchers simulate the real-time situation, but, instead of using data that are available at the time the forecasts are made, apply only a rolling or recursive estimation scheme. This procedure is strongly criticized by some researchers, namely Stark and Croushore (2002), Kozicki (2002), and Croushore (2011), who advocate the use of real-time data in forecast evaluations. This criticism can be summarized as follows:

“ If you want to analyze [...] forecasts, you must use real-time data, or your results are irrelevant; given the existence of real-time data sets for many countries, there is no excuse for not using real-time data.”

CROUSHORE (2011)

The main argument here is that models should be evaluated with the data available at the point when the forecasts are made. Although this issue is uncontroversial when econometric models are compared with professional forecasts, the consequences for model selection and evaluation are less clear.

The use of real-time data comes at a cost: it increases the number of time series to be used substantially. Furthermore, the data are easy to obtain for long-enough periods only for some countries, such as the USA or euro area. Before using real-time data, there should be clear advantages in doing so. Previously, Swanson (1996) compared bivariate Granger causality tests for main macroeconomic variables for different vintages of data and found that the test results vary significantly when different vintages are used. This finding would suggest that the data vintage may also matter in terms of out-of-sample comparisons. Denton and Kuipers (1965) investigated the consequences of data revision on the forecast accuracy of a traditional macroeconomic model and found sizable differences in performance when results on current vintage data are compared with those using the latest releases. These differences tend to be larger when only a small sample is available for comparison (Stark and Croushore, 2002) or when the data under investigation are subjected to large revisions (Kozicki, 2002).

Another issue regarding data subject to revision involves the specified forecasting target. On the one hand, a target may be able to predict the first release of GDP, which is often the case in a real-time policy environment, where the prediction error is directly measurable. Subsequent revisions are then ignored. On the other hand, one may be interested in the true state of the economy and, therefore, a later GDP vintage could be a better proxy. Choosing the specific target, thus, presents a greater impact than subsequent releases (Sestieri, 2014).

Today, more and more forecast comparisons and evaluations tend to incorporate this critique and base their analyses on fully real-time data sets (e.g., Clements and Galvão, 2009; Camacho and Perez-Quiros, 2010; Banbura, Giannone, Modugno, and Reichlin, 2013; Schorfheide and Song, 2015). In this study, we ask whether the use of real-time or currently available data produces differences when evaluating indicator models in typical nowcasting situations, i.e., forecasting the current period.

The literature on this issue is relatively unspecific. Stark and Croushore (2002) show that real-time data matter in this context. Additionally, Clements and Galvão (2009) find evidence that the forecasting accuracy of indicator models (using mixed-data sampling (MIDAS)) generally worsens when current vintages are used. However, Schumacher and Breitung (2008) and Banbura, Giannone, and Reichlin (2011) argue that the use of current or final vintages is not particularly important; these authors find that it is much more important to mimic the publication lags of the relevant indicators (ragged-edge problem). Since the results of these studies are based on dynamic factor models, the effect of data revisions may be washed out by using a composite index. Therefore, the question as to whether the effect is more prominent in leading indicator models arises.

Motivated by this question, we investigate the forecasting performance of typical indicator models for nowcasting German GDP and document the differences obtained when using final and current data releases. Our leading indicator models consist of different mixed-frequency vector autoregressive models (MF-VARs), where we can directly take into account the ragged-edge and mixed-frequency problem (Hyung and Granger, 2008; Mariano and Murasawa, 2010; Kuzin, Marcellino, and Schumacher, 2011). More specifically, we analyze the predictive power of surveys (Ifo and ZEW) relative to hard data (industrial production (IP), orders) when data revisions are taken into account. Therefore, we ask whether data revisions on IP and orders (which are substantially revised over time) present any impact on their relative importance compared with soft data, which are rarely revised.

The results show that the forecast performance of different models does not differ significantly according to the data vintage applied. On the one hand, we find that forecasts from real-time data can systematically differ from those based on latest data release, which can be explained by the existence of benchmark revisions leading to mean shifts of the entire time series. On the other hand, the relative forecasting performance of the employed models is not especially affected by the data vintage used. Although hard indicators tend to perform relatively better than surveys when the latest-available data are used, the general ranking is only marginally affected. Generally, we find that forecasts based on real-time reject the null hypothesis of equal forecast accuracy less often than those based on the latest-available vintage. This result reveals that forecasts based on

real-time data are more volatile than forecasts based on latest-available data. Moreover, we find that latest-available data are harder to predict than first releases. This finding is consistent with the news hypothesis of Mankiw and Shapiro (1986), which implies that earlier releases are revised by newly available information and these revisions cannot be predicted by past available data.

The remainder of this paper is structured as follows: We briefly characterize in Section 2 the data set used in the analysis. Then in Sections 3 and 4, we present our forecasting models and present our procedures to evaluate forecasts. We proceed with a discussion of our main results in Section 5, describe some robustness checks in Section 6, and finally provide overall conclusions in Section 7.

2 Data Set

Most of the data used in this paper are obtained from the real-time data base of the Deutsche Bundesbank, and our target variable is quarterly GDP growth adjusted by seasonal and calendar effects. Economic tendency surveys from the Ifo and ZEW are obtained from Datastream and measured on a monthly frequency. These data are only marginally revised, and we take only the latest vintage data of these variables.¹ For hard data, i.e., IP and new orders, 192 monthly data vintages from January 2000 to December 2015 are used.

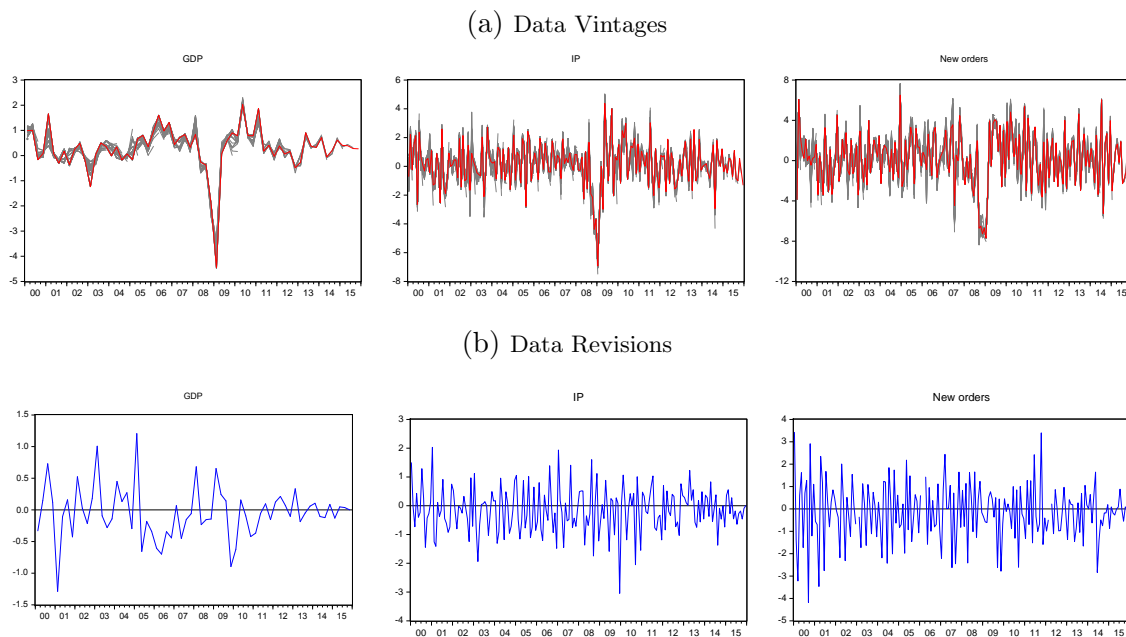
The revision process of new orders and IP was analyzed by Knetsch and Reimers (2009), who found that first releases of growth rates on understate later estimates on average. For IP, Jacobs and Sturm (2005) find some evidence that the Ifo index is able to predict future revisions of IP. Taken together, these findings suggest that first estimates of IP and new orders may be not fully optimal; instead, these estimates may be biased and/or inefficient. Faust, Rogers, and Wright (2005) report that German GDP revisions tend to be relatively large compared with those in other industrialized countries (at least until the 1990s); they also find some predictability for German GDP revisions.

Figure 1a shows different vintages of GDP growth, IP and new orders. Figure 1b shows the corresponding differences between the first and latest data releases. While the mean revision for GDP growth is nearly zero (see Table 1), that of the two monthly indicators is around -0.15 percentage points. This result is in line with previous findings by Jung (2003), who reported a mean IP revision of -0.14 percentage points and mean order revision of -0.12 percentage points. Hence, in both cases, the first release underestimates the final one. Standard deviations are also similar to those found by Jung (2003).

Another important feature of our data set is that the target variable, GDP, as well as the indicators, are released with different publication lags. Once these publication lags are taken into account in the forecasting set up, missing observations are generated at the end of the estimation sample (“ragged edges”). These publication lags must be taken into account during short-term forecasting

¹ In principle both series are revised over time, but the major source of revision is seasonal adjustment. Changes in Ifo indices over time may also occur because of the re-weighting of sub-sectors. Employing complete vintages would clearly produce more realistic results, but we are unable to do so because the real-time data set of Ifo in real-time is not available.

Figure 1: Real-time data set



Note: Panel (a) shows the growth rates to the previous period in terms of percentage of all releases in gray. The red line indicates the last available release (with all data for 2015). Panel (b) shows the revisions between the first and last (December 2015) releases of the respective period. Negative values indicate that the first release is lower than the latest-available and vice versa.

Source: German Federal Statistical Office, author's calculation.

(Giannone, Reichlin, and Small, 2008; Heinisch and Scheufele, 2017) even when using the latest-available data. For instance, German GDP is published about 6 weeks after the reference quarter but many monthly indicators are released in a more timely manner. *Hard data*, such as IP and new orders, are released relatively late (lag of about 1.5 months) compared to *soft data*, such as tendency surveys (Ifo and ZEW), which are available within the same month.

Besides timeliness, *soft data* have the additional advantage that they are hardly revised, while *hard data* – as shown above – may be subject to substantial and systematic data revisions. However, many studies have shown that once *hard data* are available, they feature strong predictive power for GDP (Banbura and Rünstler, 2011; Banbura, Giannone, Modugno, and Reichlin, 2013).

Table 1: Revision characteristics

	Mean Revision	Mean Absolute Revision	Root Mean Squared Revision
GDP	-0.045 (0.051)	0.297	0.410
Industrial production	-0.150 (0.035)	0.617	0.790
New orders	-0.142 (0.055)	1.011	1.298

Note: Sample from 2000 to 2015 from the first release to the final (December 2015) data. Standard errors are given in parenthesis.

3 Model Specifications

Leading indicator models play an important role in assessments of the current and short-term economic outlook. Typically, these models consist of bridge equation(s) that link GDP to one or more indicators (see Kitchen and Monaco, 2003, for the US). For Germany, economic tendency surveys (e.g. Ifo climate or ZEW) and new orders, all of which are observed at monthly frequency, are often considered indicators with some predictive power for short-term GDP. Most of the existing studies do not explicitly take into account the mixed-frequency nature of the data and aggregate higher frequent variables (usually leading indicators) to quarterly frequency (Hinze, 2003; Dreger and Schumacher, 2005; Kholodilin and Siliverstovs, 2006; Drechsel and Scheufele, 2012a). However, the temporal aggregation of monthly information may be associated with information loss, which could be sizable in typical nowcasting settings (see Wohlrabe, 2009b; Forni and Marcellino, 2013).

To address the mixed-frequency problem, researchers have proposed methods that explicitly take into account different sampling frequencies (see Wohlrabe, 2009a; Kuzin, Marcellino, and Schumacher, 2011; Forni and Marcellino, 2013, 2014).² From this discussion two methods that are successful in dealing with mixed frequencies have been established: MIDAS models and MF-VARs. While the former method takes into account higher frequency information via distributed lag polynomials and provides a direct forecast of the low-frequency variable, MF-VARs jointly model the variables under investigation at the higher frequency and deal with missing observations of the low-frequency variable. In the following sections, we make use of MF-VARs to analyze different leading indicators for forecasting German GDP in real-time (as proposed by Mittnik and Zadrozny, 2005; Hyung and Granger, 2008; Mariano and Murasawa, 2010; Kuzin, Marcellino, and Schumacher, 2011). The great advantage of MF-VARs is that they impose no a priori restrictions on the dynamics of the calculations.³

More specifically, the model can be represented in state-space form via the following measurement equation:

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = H_t \mathbf{s}_t, \quad (1)$$

where x_t is the quarterly GDP growth, z_t is the indicator observed at monthly frequency, and \mathbf{s}_t consists of the monthly state variables. The lag polynomial H_t is defined as

$$H_t = \begin{bmatrix} 1/3 & 0 & 2/3 & 0 & 1 & 0 & 2/3 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ for month } t = 3, 6, 9, \dots \quad (2)$$

² For Germany, only a small number of papers have explicitly taken into account the mixed-frequency problem by investigating leading indicators (Mittnik and Zadrozny, 2005; Schumacher and Breitung, 2008; Wohlrabe, 2009b; Heinisch and Scheufele, 2017).

³ See Kuzin, Marcellino, and Schumacher (2011) for a description of the pros and cons of MIDAS and MF-VARs and their application to forecasting euro area GDP. Forni, Ghysels, and Marcellino (2013) recently provided a literature review on MF-VARs and applied them to to forecast US GDP. From a theoretical perspective, Bai, Ghysels, and Wright (2013) demonstrated that MF-VARs could provide optimal forecasting properties in a mixed-frequency setup.

and

$$H_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ otherwise.} \quad (3)$$

Note that the specification of H_t implies that GDP growth x_t is only observed every third month, and that it links month-to-month growth rates x_t^* , given the aggregation rule $x_t = 1/3x_t^* + 2/3x_{t-1}^* + 3/3x_{t-2}^* + 2/3x_{t-3}^* + 1/3x_{t-4}^*$. The state equation is simply defined as

$$\mathbf{s}_t = A\mathbf{s}_{t-1} + \mathbf{u}_t, \quad (4)$$

Or, more specifically,

$$\begin{bmatrix} x_t^* \\ z_t^* \\ x_{t-1}^* \\ z_{t-1}^* \\ x_{t-2}^* \\ z_{t-2}^* \\ x_{t-3}^* \\ z_{t-3}^* \\ x_{t-4}^* \\ z_{t-4}^* \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1}^* \\ z_{t-1}^* \\ x_{t-2}^* \\ z_{t-2}^* \\ x_{t-3}^* \\ z_{t-3}^* \\ x_{t-4}^* \\ z_{t-4}^* \\ x_{t-5}^* \\ z_{t-5}^* \end{bmatrix} + \begin{bmatrix} e_t^x \\ e_t^z \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (5)$$

where $\begin{bmatrix} e_t^x \\ e_t^z \end{bmatrix} \sim N(0, \Sigma)$. Given this system, 11 parameters must be estimated (8 VAR parameters, two variances, and one covariance).⁴ Note that we do not estimate constants in the VAR; instead, we demean both time series for the estimation sample and add the mean to the forecasts afterwards. Estimation of the parameters is done by maximum likelihood.⁵ After parameter estimation, GDP growth predictions are possible with the Kalman smoother, which allows prediction irrespective of the real-time availability of GDP and the respective indicator. Those forecasts are delivered iteratively.

Our initial estimation sample covers the period 1991M2-1999M12 and expands with each new monthly release of the indicators. Similar to Schumacher and Breitung (2008), recursive estimations for the 192 vintages and, hence, 192 real-time forecasts (comprising three different forecasts for each quarter) are conducted until December 2015.

⁴ All of our results are based on a VAR(2), as suggested by Mittnik and Zdrozny (2005). Kuzin, Marcellino, and Schumacher (2011) find evidence that the lag length of their MF-VAR varies between one and two lags. For a VAR(1), the same conclusions can be drawn but the average forecast errors are slightly larger.

⁵ It is crucial to use good starting values for parameter estimation so that the estimates converge. Therefore, we choose to estimate an auxiliary regression in which apply a monthly interpolated GDP series (using a local quadratic polynomial) and the monthly indicator. From this regression, we use the starting values for the VAR parameters and the covariance matrix. Additionally, we restrict the covariance matrix to be positive definite.

To compare real-time and pseudo real-time results, we additionally run pseudo real-time regressions, where the final vintages of GDP and indicators are truncated manually and the real-time situation is mimicked as closely as possible. Similar to the real-time situation, we also take into account publication lags; thus, the availability of indicators differs between the middle and end of a month. More specifically, we take into account nine different forecasting rounds with different states of information. The first forecasting round is conducted in the middle of the first month. In this case, only information from the previous quarter is available (specifically, surveys from the last month of the previous quarter). As hard data are released later, they cannot be employed for the first two forecasting rounds. For all indicators under investigation, we look at four different cases: the last month of the previous quarter, only one month, two months, and the complete quarter (all three months).

4 Forecast Evaluation

Does a model’s forecasting performance differ when real-time data or last releases are used to conduct forecasts? To answer this question, we proceed in three steps. First, we compare forecasts from the two different approaches with each other. Second, we look at the relative performance of indicator forecasts relative to univariate benchmark forecasts. Finally, we investigate whether forecasts based on IP may be improved by the combination with other sources of information. In particular, we look at encompassing tests and investigate the difference between real-time and pseudo real-time data.

4.1 Relationship between real-time and pseudo real-time forecasts

In this section, we compare model forecasts obtained from the real-time database with those obtained from final data. Using simple measures such as plots, correlations, and moments, we provide a fundamental comparison of the two forecasting approaches. Additionally, we conduct simple tests on bias and efficiency to determine whether forecasts obtained from real-time data systematically differ from those obtained from pseudo real-time settings. More specifically, we investigate whether forecasts obtained from the latest-available vintage are good predictors of forecasts based on real-time vintages. Therefore, we look at forecast optimality tests.

First, we investigate whether the forecast means of the two approaches differ systematically, which can be tested by running the following regression:

$$\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = c + u_t, \tag{6}$$

where $\hat{x}_t^{r,i}$ is the forecast obtained with real-time data and $\hat{x}_t^{l,i}$ is the corresponding forecast based on final data.

Second, to determine whether there exists a systematic relationship between real-time forecasts and forecasts based on final data, we run a Mincer–Zarnowitz-type regression equation for efficiency (as applied by Faust, Rogers, and Wright, 2005), namely

$$\widehat{x}_t^{r,i} - \widehat{x}_t^{l,i} = \alpha + \beta \widehat{x}_t^{l,i} + e_t. \quad (7)$$

Whenever $\alpha = \beta = 0$, forecasts based on final data can be seen as optimal predictors of forecasts based on real-time data.

4.2 Relative forecasting performance

In this section, we focus on forecast accuracy relative to a benchmark model. Since absolute forecast errors are difficult to interpret, we follow common practice and compare the leading indicator forecasts based on the MF-VARs with those of univariate time-series models (Granger and Newbold, 1977; Stock and Watson, 2003).⁶ As a result, we obtain information describing by how much a leading indicator based forecast is better than the benchmark (in percentage points).

Our analysis is based on a squared loss function; thus it utilizes the same criteria for forecast evaluation as those applied for model estimation. Note that forecast errors can be calculated in different ways and typically vary with the target variable. In this study, we distinguish two cases. First, we consider the first releases of GDP as our benchmark. This benchmark type takes account of the fact that in practice many forecasts are judged by the first official estimate. Second, the latest-available vintage of GDP is used as the benchmark since this series can be seen as the best from today’s perspective and closest to *true* GDP growth.⁷

Besides the average relative performance of a model’s forecast, we evaluate whether an indicator-based forecast is systematically better than the benchmark. This evaluation must to be judged by statistical tests of equal predictive ability.⁸ Given the squared loss function, the corresponding null hypothesis of an indicator model i can be expressed as

$$H_0 : E [\Delta L_t^i] = E [(x_t - \widehat{x}_t^i)^2 - (x_t - \widehat{x}_t^{AR})^2] = 0, \quad (8)$$

where ΔL_t^i is the loss differential of the time series. One popular test for this hypothesis is the Diebold-Mariano test of equal predictive ability (Diebold and Mariano, 1995).

⁶ We use the AR(1) model as our benchmark univariate time-series model. Note that AR-forecasts differ when using real-time and pseudo real-time data.

⁷ While levels may differ due to main revisions and, hence, level shifts, growth rates remain comparable.

⁸ Typically, inference is complicated by the fact that forecasts are generated by estimated models and depend on parameter uncertainty (West, 1996). However, since our models under investigation are non-nested, the issue of parameter estimation uncertainty can be ignored in our setting (West, 2006). The issue of employing real-time data has been analyzed by Clark and McCracken (2009). In general, tests should be valid in the case of unpredictable revisions (and non-nested models).

Instead of relying on parametric tests such as the Diebold-Mariano-West (DMW) test, we apply nonparametric tests based on signed-rank statistics. A clear advantage of these tests is that they are invariant to deviations from standard assumptions, such as those of normality or homoscedasticity (Lehmann, 1975). The applied Wilcoxon signed-rank test is robust to (conditional) heteroscedasticity, heavy tails, and outliers in the loss-differential series, which is particularly important when analyzing samples including specific events such as financial crises where loss differences are subject to unusual behavior (large outliers and heteroscedasticity).⁹ Nevertheless, these tests present very good power characteristics compared with the DMW test (Luger, 2004).

The test statistic is given by

$$W = \sum_{t=1}^T u(\Delta L_t) \text{rank}(|\Delta L_t|), \quad (9)$$

where $u(\Delta L_t) = 1$ if the loss differential is positive ($\Delta L_t > 0$) and 0 otherwise. Critical values for W are tabulated and, for $T > 20$, an approximation by normality exists. This test has been discussed and applied in similar settings by Campbell and Dufour (1995), Campbell and Ghysels (1995), and Diebold and Mariano (1995).

One shortcoming of the basic test, however, is that it requires the absence of serial correlation in the loss differential. As we also consider multi-period forecasts (e.g., when GDP of the previous quarter is not available), this assumption may not be satisfied. Therefore, we follow the suggestion of Campbell and Dufour (1995) and apply split-sample techniques. We divide our original dependent loss differential into k independent ones. For instance, in the case of MA(1) in the loss differential we would define two sample sets (A and B) consisting of observations $T_A = 1, 3, 5, \dots$ and $T_B = 2, 4, 6, \dots$. Then, the signed-rank test is applied on each of the two subsamples A and B . The null hypothesis can be rejected if either W_A or W_B is significant at the $\alpha/2$ level, which implies that we apply a Bonferroni correction for multiple comparisons. To decide whether the split-sample technique is applicable, we use the test of serial dependence by Dufour (1981), which is also based on the signed-rank principle and shares the similar robustness properties.¹⁰

4.3 Benefits of using additional information sources

Finally, we check whether GDP nowcasts based on hard data (i.e., IP) can be improved by combining them with other information sources. This endeavor is related to the issue of whether combining two models results in higher forecast accuracy. Bates and Granger (1969), Clemen (1989), and Timmermann (2006) provide theoretical and empirical evidence that model combination results in higher forecast accuracy. In our case this issue is investigated by means of forecast encompassing

⁹ By looking at our loss differences ΔL_t we see very substantial deviations from normality, specifically, very large excess kurtosis and outliers.

¹⁰ We consider serial correlation up to order 3 ($k = 3$), which translates into a maximum of four subsamples. In our empirical application, we find mostly evidence of $k = 0$ and $k = 1$.

(Chong and Hendry, 1986), and we employ Harvey, Leybourne, and Newbold (1998)'s forecast encompassing test.¹¹ This test is based on the regression

$$\widehat{e}_t^{ip} = \lambda \left(\widehat{e}_t^{ip} - \widehat{e}_t^i \right) + v_t \quad (10)$$

for each indicator i . The corresponding null hypothesis equals $H_0 : \lambda = 0$, where λ is the optimal weight of the combined forecast of indicator i with the ip forecast. Whenever the test is rejected, the combined forecast is superior to the single IP-based model. In this case, our particular interest lies in whether the optimal combination weights differ when real-time data are used.

¹¹ To achieve the encompassing tests, we opted for standard parametric tests and did not use nonparametric procedures since no strong evidence of violation of the basic assumptions was observed. Note that these tests can be formulated in terms of forecast errors instead of squared errors, which is the case in comparisons of forecast accuracy.

Table 2: Forecast vintage comparisons

	all months	two months	one month	no month
I. Industrial production				
a. Descriptive analysis:				
Spearman rank corr	0.719	0.741	0.774	0.711
Var ratio $\frac{\sigma_r^2}{\sigma_l^2}$	1.051	1.043	1.223	0.905
b. Bias test:				
c (mean difference)	-0.133	-0.070	-0.057	-0.098
p-val	0.001	0.040	0.041	0.007
c. Efficiency test				
α	-0.108	-0.032	-0.046	0.001
β	-0.061	-0.096	-0.032	-0.277
p-val ($\alpha = \beta = 0$)	0.099	0.138	0.644	0.000
II. New orders				
a. Descriptive analysis:				
Spearman rank corr	0.850	0.888	0.854	0.827
Var ratio $\frac{\sigma_r^2}{\sigma_l^2}$	0.955	1.010	1.107	1.145
c. Bias test:				
c (mean difference)	-0.043	-0.009	-0.057	-0.079
p-val	0.102	0.679	0.049	0.011
b. Efficiency test:				
α	-0.016	0.010	-0.044	-0.050
β	-0.074	-0.052	-0.035	-0.083
p-val ($\alpha = \beta = 0$)	0.058	0.178	0.508	0.303
III. Ifo situation				
a. Descriptive analysis:				
Spearman rank corr	0.977	0.976	0.947	0.965
Var ratio $\frac{\sigma_r^2}{\sigma_l^2}$	0.848	0.860	0.744	0.737
b. Bias test:				
c (mean difference)	-0.031	-0.028	-0.016	-0.019
p-val	0.062	0.085	0.558	0.420
c. Efficiency test:				
α	0.009	0.010	0.058	0.049
β	-0.089	-0.084	-0.171	-0.165
p-val ($\alpha = \beta = 0$)	0.000	0.000	0.000	0.000
IV. Ifo expectations				
a. Descriptive analysis:				
Spearman rank corr	0.943	0.949	0.980	0.950
Var ratio $\frac{\sigma_r^2}{\sigma_l^2}$	0.887	0.864	0.843	0.845
b. Bias test:				
c (mean difference)	-0.029	-0.017	-0.011	-0.006
p-val	0.072	0.311	0.352	0.684
c. Efficiency test:				
α	0.012	0.024	0.028	0.036
β	-0.095	-0.096	-0.094	-0.101
p-val ($\alpha = \beta = 0$)	0.000	0.002	0.002	0.000

Note: Forecasts obtained with real-time data are compared with those obtained using final data (available for December 2015). The bias test is based on $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = c + u_t$. The efficiency test is based on $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = \alpha + \beta \hat{x}_t^{l,i} + e_t$.

5 Results

In this section, we present our major results. First, we show how forecasts differ when models are estimated with real-time or final data. Second, we look at the relative performance of the indicator models. Third, we investigate how encompassing tests are affected by using real-time or latest-available data.

5.1 Forecast differences

Figure 2 presents a descriptive method to analyze the association between the two forecast procedures. Table 2 shows a more systematic approach to investigate the relation between real-time and pseudo real-time forecasts. These two ways of looking at the results suggests a very high correlation between forecasts obtained with real-time and final data. The Spearman rank correlation is quite high and ranges from 0.71 (IP) to close to 1 in the case of the surveys.

To determine whether a systematic difference exists between the two ways of conducting forecasts, we test for bias and efficiency. For IP and new orders, we find clear evidence of the systematic downward bias of real-time forecasts when compared with those obtained from final-released data. This result implies that GDP growth forecasts based on real-time data are, on average, lower than those obtained from final data. Two reasons may explain this finding. First, the downward bias in early IP and new order figures (see Section 2) may cause a lower GDP growth nowcast. Second, the higher mean in GDP growth of final estimates may also translate into higher average forecasts.

In terms of surveys, we find a different pattern. The efficiency tests are always significant, which indicates that is the absence of a one-to-one relation between real-time and final data forecasts. Instead, GDP forecasts based on final-available data are slightly more volatile than those based on real-time data. Therefore, one has to multiply the latest-available forecast by a factor of 0.83–0.91 to obtain the real-time forecast. In the next subsection, we will discuss whether these forecast differences present practical implications for model comparisons.

5.2 Relative forecasting performance

Our analysis features several dimensions. First, we differentiate between two different target variables: (a) the last-available GDP vintage and (b) the first-available GDP releases. Second, we distinguish two approaches of model estimation and specify from which data vintage the forecasts are conducted. In this dimension, we obtain forecasts based on (i) real-time vintage data (i.e., real-time forecasts) and on (ii) final data (latest-available vintage, i.e. pseudo real-time forecasts). A third dimension is the type of indicator used to predict GDP. Here we consider IP and new orders as *hard* indicators and tendency surveys from Ifo and ZEW as *soft* indicators.

Forecast target: Latest GDP

Table 3 and Figures 3 and 4 show the forecasting performance results obtained using real-time and final data to predict final GDP growth.¹² In general, the results confirm findings in previous studies that additional information tends to improve forecast accuracy (as shown by Banbura, Giannone, Modugno, and Reichlin, 2013; Heinisch and Scheufele, 2017, for typical nowcasting applications). This result is particularly true for hard data. Most indicators show higher forecast accuracy than the univariate benchmark. Exceptions to this trend are the ZEW indicators where improvements are very small or do not exist. In many cases, differences compared with the autoregressive (AR) model are not significant.

Table 3: Relative RMSFEs (Benchmark: Final release of GDP growth)

	M1 mid	M1 end	M2 mid	M2 end	M3 mid	M3 end	M4 mid	M4 end	M5 mid
Forecasts based on real-time data									
AR	0.895	0.895	0.895	0.878	0.878	0.878	0.878	0.878	0.878
IP	NA	NA	0.870	0.887	0.700	0.700	0.720	0.720	0.601 **
New orders	NA	NA	0.870	0.887	0.764	0.764	0.730	0.730	0.732
Ifo situation	0.820	0.829	0.829	0.778	0.778	0.768 *	0.768 *	0.768 *	0.768 *
Ifo expectation	0.806 *	0.798 **	0.798 **	0.826 *	0.826 *	0.811 **	0.811 **	0.811 **	0.811 **
Ifo climate	0.841	0.785	0.785	0.787	0.787	0.794	0.794	0.794	0.794
ZEW situation	0.995	0.965	0.965	0.983	0.983	0.984	0.984	0.984	0.984
ZEW expectation	0.976	0.961	0.961	1.022	1.022	0.988	0.988	0.988	0.988
Forecasts based on latest available data									
AR	0.903	0.903	0.903	0.879	0.879	0.879	0.879	0.879	0.879
IP	NA	NA	0.834 *	0.858	0.686 **	0.686 **	0.601 *	0.601 *	0.613 **
New orders	NA	NA	0.759	0.780	0.729 *	0.729 *	0.724	0.724	0.727
Ifo situation	0.804	0.805	0.805	0.784	0.784	0.771	0.771	0.771	0.771
Ifo expectation	0.780	0.784 *	0.784 *	0.812 **	0.812 **	0.794 **	0.794 **	0.794 **	0.794 **
Ifo climate	0.798	0.779	0.779	0.780	0.780	0.766	0.766	0.766	0.766
ZEW situation	0.976	0.940	0.940	0.963	0.963	0.967	0.967	0.967	0.967
ZEW expectation	0.973	0.963	0.963	1.114 *	1.114 *	1.025	1.025	1.025	1.025

Note: Real-time and pseudo real-time forecasts compared with **last GDP releases** for December 2015. Relative root mean squared forecast errors (RMSFEs) are given for indicator forecasts compared with those of the univariate forecasts. Tests on equal forecast ability are based on signed-rank tests. *, ** indicates significance at the 10% and 5% levels, respectively.

Hard indicators are more advantageous than soft indicators when one month of the respective quarter is available, which is the case in the middle of the third month of a quarter. In this case, hard data can improve benchmark forecasts by up to 30%–40%. Surveys dominate earlier in the quarter, but offer only marginal improvements when additional data are available. Forecasting gains based on the most promising survey data (Ifo climate and expectations) are around 20%.

When differences between real-time or final data used for estimation are compared, we see that hard data forecast accuracy is slightly better than other results when it is based on final data. The overall ranking of different indicators given the specific state of information is not significantly affected.

However, slight differences in the tests on equal forecasting ability are observed. These tests reject the null hypothesis of equal forecasting ability more often in the case of final data than in the

¹² See Figure 6 in the Appendix.

Table 4: Relative RMSFEs (Benchmark: First release of GDP growth)

	M1 mid	M1 end	M2 mid	M2 end	M3 mid	M3 end	M4 mid	M4 end	M5 mid
Forecasts based on real-time data									
AR	0.792	0.792	0.792	0.788	0.788	0.788	0.788	0.788	0.788
IP	NA	NA	0.837	0.841	0.616	0.616	0.609	0.609	0.479 **
New orders	NA	NA	0.803	0.808	0.659	0.659	0.649	0.649	0.661
Ifo situation	0.843	0.841	0.841	0.790	0.790	0.787	0.787	0.787	0.787
Ifo expectation	0.832	0.817	0.817	0.845 *	0.845 *	0.829 *	0.829 *	0.829 *	0.829 *
Ifo climate	0.884	0.819	0.819	0.813	0.813	0.835	0.835	0.835	0.835
ZEW situation	1.027	0.988	0.988	1.005	1.005	1.002	1.002	1.002	1.002
ZEW expectation	0.974	0.954	0.954	1.027	1.027	0.978	0.978	0.978	0.978
Forecasts based on latest available data									
AR	0.799	0.799	0.799	0.791	0.791	0.791	0.791	0.791	0.791
IP	NA	NA	0.859	0.867	0.676 *	0.676 *	0.521 **	0.521 **	0.513 **
New orders	NA	NA	0.697 *	0.703 *	0.620 **	0.620 **	0.642	0.642	0.656
Ifo situation	0.838	0.827	0.827	0.802	0.802	0.799	0.799	0.799	0.799
Ifo expectation	0.793	0.792	0.792	0.818 **	0.818 **	0.800 *	0.800 *	0.800 *	0.800 *
Ifo climate	0.845	0.807	0.807	0.789	0.789	0.778	0.778	0.778	0.778
ZEW situation	1.005	0.961	0.961	0.978	0.978	0.978	0.978	0.978	0.978
ZEW expectation	0.978	0.950	0.950	1.129	1.129	1.033	1.033	1.033	1.033

Note: Real-time and pseudo real-time forecasts compared with **first GDP releases**. Tests on equal forecast ability are based on signed-rank tests. *, ** indicates significance at the 10% and 5% levels, respectively.

case of real-time data. Ifo expectations is the only indicator that does significantly better than the benchmark in most of the cases studied. One reason that most of the tests are not significant suggests that stability might be a problem (although average improvements are quite sizable).

Forecast target: First GDP

When the first GDP release is used as the benchmark for forecasting comparisons, the overall results are very similar (Table 4). In terms of relative gains in forecast accuracy and significance, the results differ only slightly.

However, the relative gains for hard data are slightly larger when compared with first GDP releases. This finding suggests that the statistical office relies on IP and new orders as inputs for early GDP estimates when other information is relatively scarce.

Another interesting finding from using first GDP releases as the forecast target is that first GDP releases are easier to predict, as reflected by the lower RMSFEs of the benchmark AR model. This result implies that first GDP releases are less volatile than later releases, consistent with the news hypothesis (Mankiw and Shapiro, 1986), which states that first GDP releases are efficient predictors of later releases.

5.3 Combination of different sources of information

To assess whether the forecast accuracy improves when combining the IP forecast with other information sources, we conducted encompassing tests. Table 5 indicates that the orders and the

Table 5: Encompassing tests

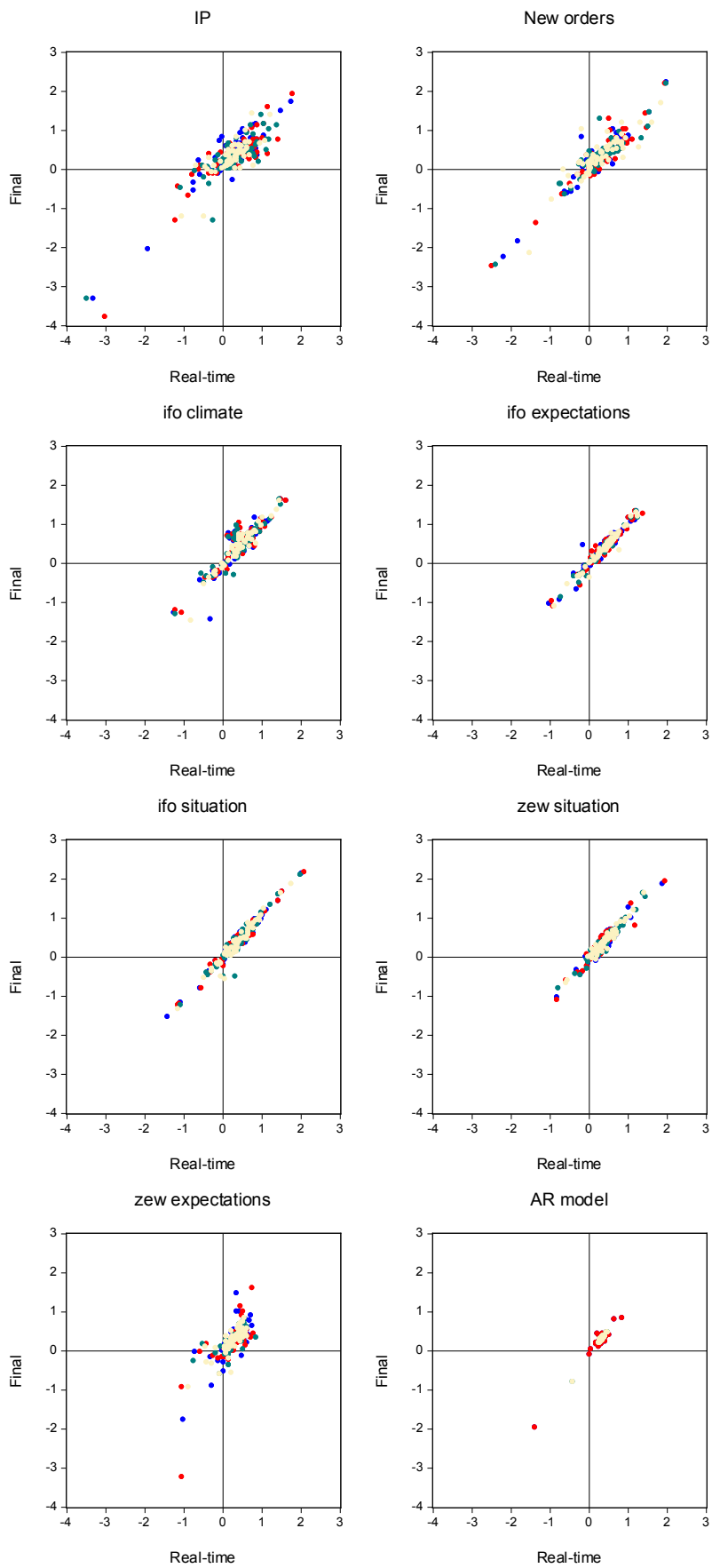
	M2 mid	M2 end	M3 mid	M3 end	M4 mid
Forecasts based on real-time data					
New orders	0.50 **	0.50 **	0.31 *	0.31 *	0.47 ***
Ifo situation	0.60 ***	0.76 ***	0.38 **	0.40 **	0.42 **
Ifo expectation	0.73 ***	0.76 ***	0.27	0.30	0.36 *
Ifo climate	0.69 ***	0.76 ***	0.35 **	0.34 *	0.39 **
ZEW situation	0.19	0.17	0.15	0.15	0.20
ZEW expectation	0.01	-0.08	0.05	0.07	0.15
Forecasts based on latest available data					
New orders	0.89 **	0.89 **	0.32 *	0.32 *	-0.05
Ifo situation	0.60 ***	0.69 ***	0.29 *	0.33 **	0.22
Ifo expectation	0.73 ***	0.78 ***	0.13	0.19	0.14
Ifo climate	0.69 ***	0.75 ***	0.29 *	0.33 **	0.20
ZEW situation	0.19	0.08	0.01	0.02	0.04
ZEW expectation	0.01	-0.30	-0.20	-0.15	0.00

Note: The test-statistic indicates whether IP already encompasses all information or whether additional information is given by other indicators. *** indicates significance at the 1% level.

Ifo indicators offer significant improvements once they are combined with IP. Even after some months of IP are known, surveys provide some added value. The weights tend to be somewhat higher for surveys once real-time data are used.

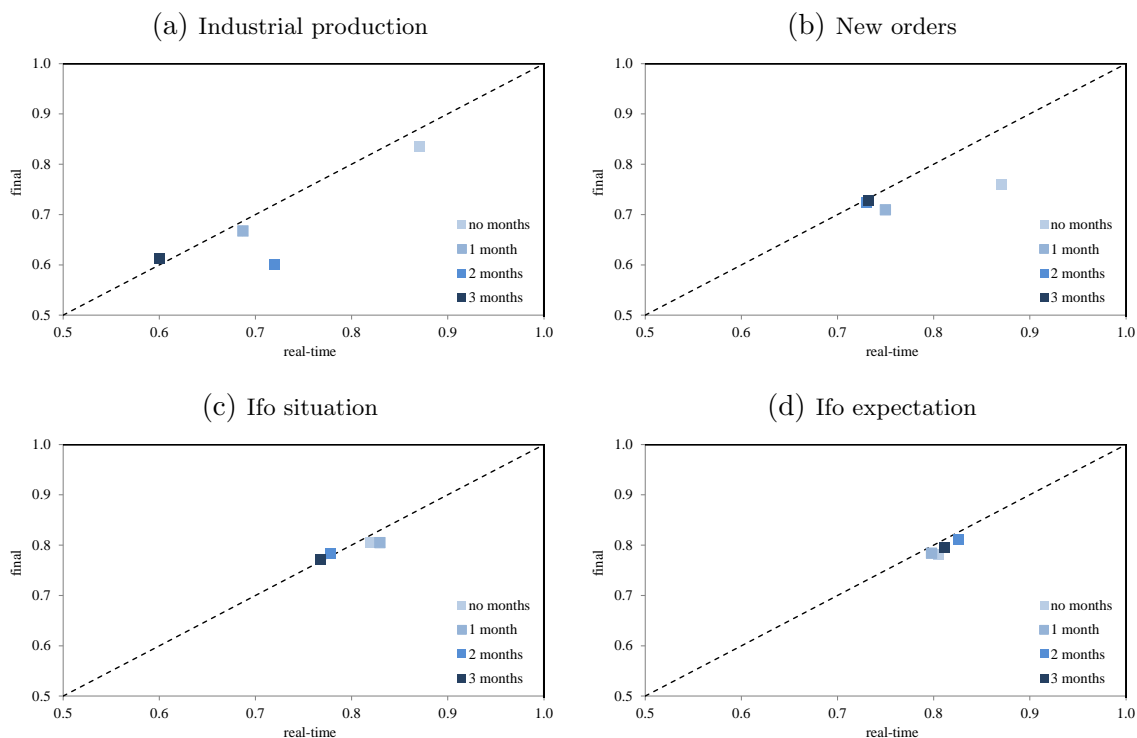
This result implies that survey data present important value as an additional data source. In the presence of data revisions, the weight of surveys may be even higher than that indicated in the case of pseudo out-of-sample comparisons. Thus, forecast combination might be beneficial in a real-time forecast setting.

Figure 2: Forecast comparison



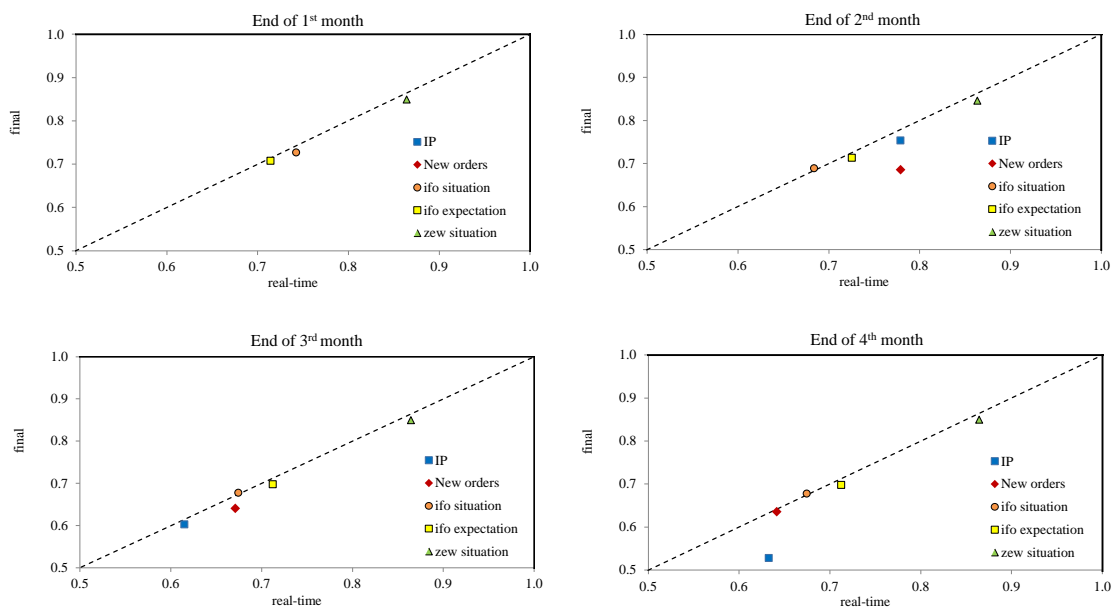
Note: Forecasts based on real-time and final data using different states of information: blue (no month), red (1 month), green (2 months), and yellow (all 3 months).

Figure 3: Forecast performance at different states of information



Note: Forecasts with different monthly information sets based on real-time and final data (compared with final-available releases for December 2015).

Figure 4: Forecast performance of different indicators



Note: End-of-month forecasts based on real-time and final data compared with final-available releases for December 2015.

6 Robustness

To check the robustness of our results we consider two additional tests. First, we investigate how our results change when forecasts are conducted using only first GDP releases for estimation. Second, we check how stable the results are over time.

6.1 Models estimated based on first releases

Koenig, Dolmas, and Piger (2003) argue that one should include only first-available estimates for the target variable in the forecasting equation and use current vintages as indicator variables to obtain reasonably good forecasts for first official GDP releases. This setting avoids the problem where the values of the targeted time series correspond to different revision stages. As an indicator, the authors suggest including only data that have been available to the forecaster at this point in time. They argue that the initial release is an efficient estimate of subsequent releases, meaning revisions are unpredictable using data at the time of the release. Table 6 shows that the relative RMSFEs are only marginally smaller compared with the ones given in Table 4.

Table 6: RMSFEs using first GDP releases for estimation

	M1 mid	M1 end	M2 mid	M2 end	M3 mid	M3 end	M4 mid	M4 end	M5 mid
Forecasts based on first GDP releases and real-time indicator data									
AR	0.788	0.788	0.788	0.772	0.772	0.772	0.772	0.772	0.772
IP	NA	NA	0.829	0.847	0.683	0.683	0.635	0.635	0.578
New orders	NA	NA	0.744	0.760	0.638	0.638	0.666	0.666	0.634
Ifo situation	0.839	0.840	0.840	0.834	0.834	0.833	0.833	0.833	0.833
Ifo expectation	0.827	0.806	0.806	0.846 *	0.846 *	0.829	0.829	0.829	0.829
Ifo climate	0.830	0.804	0.804	0.837	0.837	0.816	0.816	0.816	0.816
ZEW situation	1.018	0.969	0.969	1.011	1.011	1.009	1.009	1.009	1.009
ZEW expectation	0.990	0.961	0.961	0.995	0.995	1.035	1.035	1.035	1.035

Note: Relative RMSFEs for GDP forecasts based on various indicators are shown at different stages of information (relative to the RMSFE of the AR forecast given in the first line and shaded in gray). RMSFEs are calculated based on first GDP releases.

The results in Table 6 indicate that using only first GDP releases as the dependent variable hardly changes the earlier conclusions, where models were estimated based on real-time vintages. All of our results show that different variants of the dependent variable present little impact on the relative forecasting performance of the models.

6.2 Stability over time

The financial crisis exerted a large impact on forecast error comparisons (Drechsel and Scheufele, 2012b). Therefore, instead of looking only at the average performance over time, we take into account potential instabilities of forecasting performance.¹³ More precisely, we follow Giacomini and Rossi's (2010) fluctuation test. This test is based on rolling regression of the Diebold-Mariano

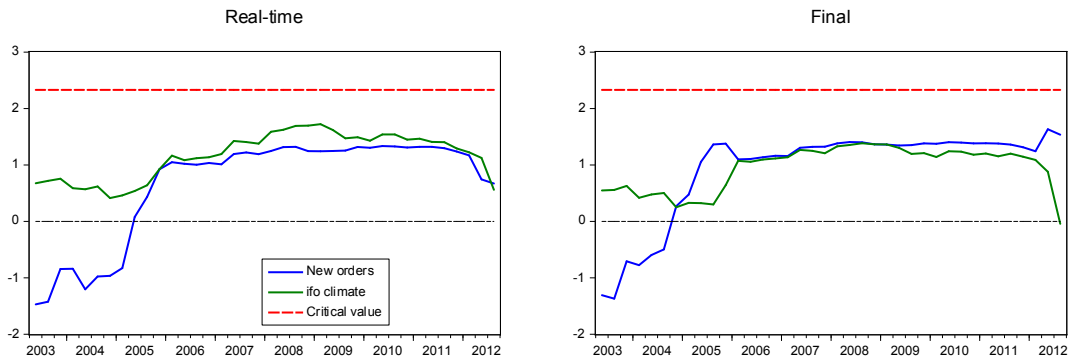
¹³ See Figures 7 and 8 in the Appendix.

test. However, as stated earlier, this test may not be very reliable because our loss differential is far from normality; thus, we do not over-interpret this test. Since the test statistic is basically a weighted mean difference, we can look at the relative performance of the indicator based forecast over time.

Figure 5 shows the test statistic for different information stages (a–c) and different indicator combinations. Positive (negative) differences indicate that the indicator model produces better (worse) forecasts than the AR model. Interestingly, we see that hard data are not very informative at the beginning of our sample. All three panels show that indicator forecasts based on hard data are inferior to AR forecasts. In 2003–2005, new orders and IP do not offer any improvements for predicting GDP (confirming the results of Kholodilin and Siliverstovs, 2006, who found deterioration of leading indicator predictability from 2001 to 2004). This result rationalizes previous findings that the forecasting performance obtained using hard data is not always significantly better than that of univariate models. In 2005, the relative forecasting performance using hard data is notably improved, but no episodes where single indicators offer very significant improvements are observed. This finding is in line with the results on average performance. Overall, the forecasting performances of the models over time using real-time or final data are very similar.

Figure 5: Local relative performance based on fluctuation tests

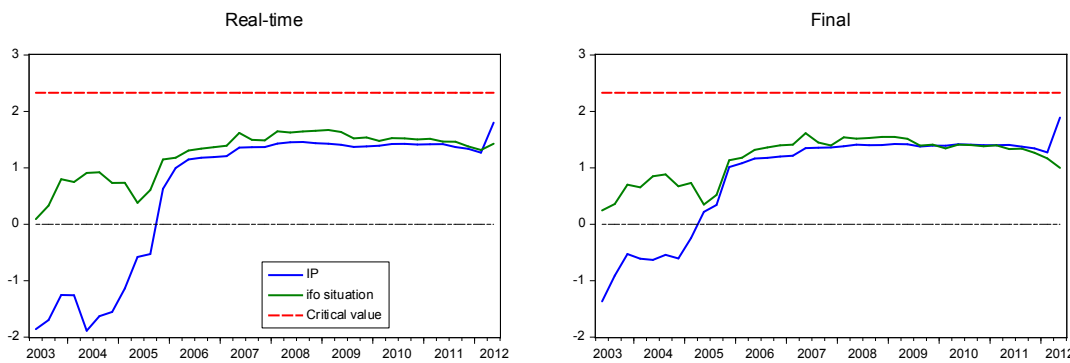
(a) End of the 2nd month



(b) End of the 3rd month



(c) End of the 4th month



Note: Local relative performance as the out-of-sample MSFE difference computed over rolling windows (the local relative MSFE). The critical value for the one-sided test is 2.33 for the 10% significance level (Giacomini and Rossi, 2010) for a rolling window of 25 quarters.

7 Conclusion

In this paper, we analyzed whether GDP forecasts differ when using real-time or latest-available data. Using an MF-VAR methodology, we take into account the mixed-frequency and ragged-edge problems in a realistic nowcasting setting for Germany.

We find that real-time and pseudo real-time forecasts differ in certain aspects. For hard data evaluations, GDP forecasts in real-time are systematically lower than the corresponding forecasts based on latest-available data. This result may be explained by the systematic downward bias of first indicator releases or by benchmark revisions that shift the mean GDP growth up. For survey evaluations, real-time GDP forecasts are less variable than forecasts based on pseudo real-time data. This result is an indirect consequence of the news hypothesis, which implies that early GDP releases are less volatile than later releases. The understated variability of first releases seems to transmit via the MF-VAR model to short-term forecasts.

Interestingly, the different characteristics of the forecasts are not significantly affected by the relative forecasting performance of different indicators. As expected, the pseudo real-time setting slightly overstates the forecasting accuracy of hard indicators. However, the differences observed are very small, and the rankings of different indicator forecasts do not change whether real-time or pseudo real-time data are used in forecasting comparisons. Much more important is to take into account the specific timing of data releases (confirming findings by Banbura, Giannone, Modugno, and Reichlin, 2013). These results imply that using real-time data is unnecessary when leading indicators are ranked according to their relative nowcasting performance. However, while the ranking of indicators is unaffected, we find some evidence that ignorance of data revisions in forecast comparisons may lead to overrejections of tests on average forecasting accuracy. Additionally, latest-available data releases may understate the role of surveys in improving forecasts in combination with hard data.

One extension of the presented analysis is the inclusion of other forecasting methods dealing with mixed-frequency data, namely, MIDAS models. Additionally, one could investigate financial variables as an additional source of data that are not revised. Obviously, the impact of data revisions on forecasting performance may be affected by the properties of the revision process, as well as by the quality of the survey data. Therefore, our results may not be directly applicable to other countries.

References

- BAI, J., E. GHYSELS, AND J. H. WRIGHT (2013): “State Space Models and MIDAS Regressions,” *Econometric Reviews*, 32, 779–813.
- BANBURA, M., D. GIANNONE, M. MODUGNO, AND L. REICHLIN (2013): “Now-Casting and the Real-Time Data Flow,” in *Handbook of Economic Forecasting*, ed. by G. Elliot, and A. Timmermann, vol. 2A, chap. 4, pp. 195–237. Elsevier.
- BANBURA, M., D. GIANNONE, AND L. REICHLIN (2011): “Nowcasting,” in *The Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements, and D. F. Hendry, chap. 7, pp. 193–224. Oxford University Press.
- BANBURA, M., AND G. RÜNSTLER (2011): “A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP,” *International Journal of Forecasting*, 27(2), 333–346.
- BANERJEE, A., M. MARCELLINO, AND I. MASTEN (2005): “Leading Indicators for Euro-area Inflation and GDP Growth,” *Oxford Bulletin of Economics and Statistics*, 67(s1), 785–813.
- BATES, J. M., AND C. W. J. GRANGER (1969): “The Combination of Forecasts,” *Operations Research Quarterly*, 20(4), 451–468.
- CAMACHO, M., AND G. PEREZ-QUIROS (2010): “Introducing the Euro-Sting: Short-term Indicator of Euro Area Growth,” *Journal of Applied Econometrics*, 25(4), 663–694.
- CAMPBELL, B., AND J.-M. DUFOUR (1995): “Exact Nonparametric Orthogonality and Random Walk Tests,” *The Review of Economics and Statistics*, 77(1), 1–16.
- CAMPBELL, B., AND E. GHYSELS (1995): “Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency,” *The Review of Economics and Statistics*, 77(1), 17–31.
- CHONG, Y. Y., AND D. F. HENDRY (1986): “Econometric Evaluation of Linear Macro-Economic Models,” *The Review of Economic Studies*, 53(4), 671–690.
- CLARK, T., AND M. MCCracken (2009): “Tests of equal predictive ability with real-time data,” *Journal of Business and Economic Statistics*, 27(4), 441–454.
- CLEMEN, R. T. (1989): “Combining forecasts: A review and annotated bibliography,” *International Journal of Forecasting*, 5(4), 559–583.
- CLEMENTS, M., AND A. GALVÃO (2009): “Forecasting US output growth using Leading Indicators: An appraisal using MIDAS models,” *Journal of Applied Econometrics*, 24(7), 1187–1206.
- CROUSHORE, D. (2011): “Frontiers of Real-Time Data Analysis,” *Journal of Economic Literature*, 49(1), 72–100.
- DENTON, F. T., AND J. KUIPERS (1965): “The Effect of Measurement Errors on Parameter Estimates and Forecasts: A Case Study Based on the Canadian Preliminary National Accounts,” *The Review of Economics and Statistics*, 47(2), 198–206.

- DIEBOLD, F. X., AND R. S. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13(3), 253–263.
- DRECHSEL, K., AND R. SCHEUFELE (2012a): “The Financial Crisis from a Forecaster’s Perspective,” *Kredit und Kapital*, 2012(1), 1–26.
- (2012b): “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.
- 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.
- DUFOUR, J.-M. (1981): “Rank Tests for series dependence,” *Journal of Time Series Analysis*, 2(3), 117–128.
- FAUST, J., J. H. ROGERS, AND J. H. WRIGHT (2005): “News and Noise in G-7 GDP Announcements,” *Journal of Money, Credit and Banking*, 37(3), 403–419.
- FORONI, C., E. GHYSELS, AND M. MARCELLINO (2013): “Mixed-frequency vector autoregressive models,” *Advances in Econometrics*, 32, 247–272.
- FORONI, C., AND M. MARCELLINO (2014): “A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates,” *International Journal of Forecasting*, 30(3), 554–568.
- FORONI, C., AND M. G. MARCELLINO (2013): “A survey of econometric methods for mixed-frequency data,” Discussion Paper 06, Norges Bank.
- GIACOMINI, R., AND B. ROSSI (2010): “Forecast comparisons in unstable environments,” *Journal of Applied Econometrics*, 25(4), 595–620.
- GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): “Nowcasting: The real-time informational content of macroeconomic data,” *Journal of Monetary Economics*, 55(4), 665–676.
- GRANGER, C. W., AND P. NEWBOLD (1977): *Forecasting Economic Time Series*. Academic Press, New York.
- HARVEY, D. I., S. J. LEYBOURNE, AND P. NEWBOLD (1998): “Tests for Forecast Encompassing,” *Journal of Business & Economic Statistics*, 16(2), 254–259.
- HEINISCH, K., AND R. SCHEUFELE (2017): “Bottom-up or Direct? Forecasting German GDP in a Data-rich Environment,” *Empirical Economics*, forthcoming.
- HINZE, J. (2003): “Prognoseleistung von Frühindikatoren: Die Bedeutung von Frühindikatoren für Konjunkturprognosen - Eine Analyse für Deutschland,” Discussion Paper Series 26253, Hamburg Institute of International Economics.
- HYUNG, N., AND C. W. J. GRANGER (2008): “Linking Series Generated at Different Frequencies,” *Journal of Forecasting*, 27(2), 95–108.

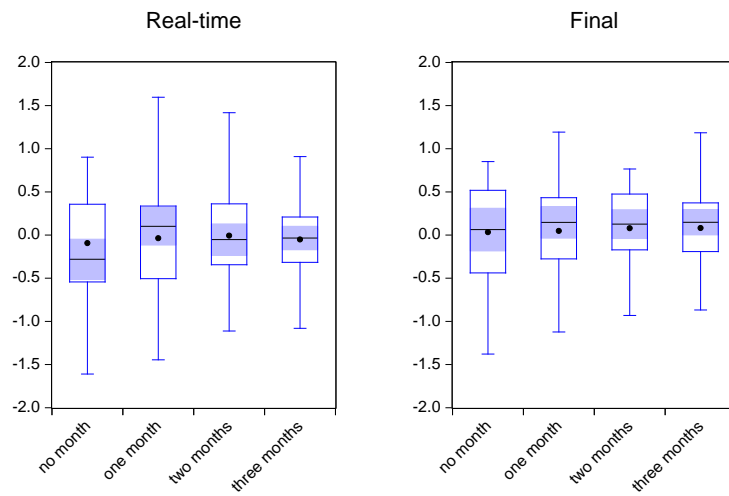
- JACOBS, J., AND J.-E. STURM (2005): “Do Ifo Indicators Help Explain Revisions in German Industrial Production?,” in *Ifo Survey Data in Business Cycle and Monetary Policy Analysis*, ed. by J.-E. Sturm, and T. Wollmershäuser, pp. 93–114. Physica.
- JUNG, S. (2003): “Revisionsanalyse des deutschen Produktionsindex,” *Wirtschaft und Statistik*, 9/2003, 819–826.
- KHOLODILIN, K. A., AND B. SILIVERSTOVTS (2006): “On the Forecasting Properties of the Alternative Leading Indicators for the German GDP: Recent Evidence,” *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, 226(3), 234–259.
- KITCHEN, J., AND R. MONACO (2003): “Real-Time Forecasting in Practice,” *Business Economics*, 38(4), 10–19.
- KNETSCH, T. A., AND H.-E. REIMERS (2009): “Dealing with Benchmark Revisions in Real-Time Data: The Case of German Production and Orders Statistics,” *Oxford Bulletin of Economics and Statistics*, 71(2), 209–235.
- KOENIG, E. F., S. DOLMAS, AND J. PIGER (2003): “The Use and Abuse of Real-Time Data in Economic Forecasting,” *The Review of Economics and Statistics*, 85, 618–628.
- KOZICKI, S. (2002): “Comments on Forecasting with a real-time data set for macroeconomists,” *Journal of Macroeconomics*, 24(4), 541–557.
- KUZIN, V., M. MARCELLINO, AND C. SCHUMACHER (2011): “MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area,” *International Journal of Forecasting*, 27(2), 529–542.
- LEHMANN, E. L. (1975): *Nonparametrics: Statistical Methods based on Ranks*. Holden-Day, INC., San Francisco.
- LUGER, R. (2004): “Exact Tests of Equal Forecast Accuracy with an Application to the Term Structure of Interest Rates,” Staff Working Papers 04-2, Bank of Canada.
- MANKIW, N., AND M. SHAPIRO (1986): “News or noise: an analysis of GDP revisions,” *Survey of Current Business*, 66(5), 20–25.
- MARIANO, R., AND Y. MURASAWA (2010): “A Coincident Index, Common Factors, and Monthly Real GDP,” *Oxford Bulletin of Economics and Statistics*, 72(1), 27–46.
- MITTNIK, S., AND P. A. ZADROZNY (2005): “Forecasting German GDP at Monthly Frequency Using Monthly ifo Business Conditions Data,” in *ifo Survey Data in Business Cycle and Monetary Policy Analysis*. Springer-Verlag.
- SCHORFHEIDE, F., AND D. SONG (2015): “Real-time forecasting with a mixed-frequency VAR,” *Journal of Business & Economic Statistics*, 33(3), 366–380.
- SCHUMACHER, C., AND J. BREITUNG (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.

- SESTIERI, G. (2014): “Comments on ”A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates”,” *International Journal of Forecasting*, 30(3), 569 – 571.
- STARK, T., AND D. CROUSHORE (2002): “Forecasting with a real-time data set for macroeconomists,” *Journal of Macroeconomics*, 24(4), 507–531.
- STOCK, J. H., AND M. W. WATSON (2003): “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature*, 47(3), 788–829.
- SWANSON, N. R. (1996): “Forecasting using first-available versus fully revised economic time-series data,” *Studies in Nonlinear Dynamics and Econometrics*, 1(1), 4764.
- TIMMERMANN, A. (2006): “Forecast Combinations,” in *Handbook of Forecasting*, ed. by G. Elliot, C. W. Granger, and A. Timmermann, vol. 1, chap. 4, pp. 135–196. Elsevier.
- WEST, K. D. (1996): “Inference about Predictive Ability,” *Econometrica*, 64(5), 1067–1084.
- (2006): “Forecast Evaluation,” in *Handbook of Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann, vol. 1, chap. 3, pp. 99–134. Elsevier, Amsterdam.
- WOHLRABE, K. (2009a): “Forecasting with Mixed-frequency Time Series Models,” Ph.d. thesis, University Munich.
- (2009b): “Makroökonomische Prognosen mit gemischten Frequenzen,” *ifo Schnelldienst*, 62(21), 22–33.

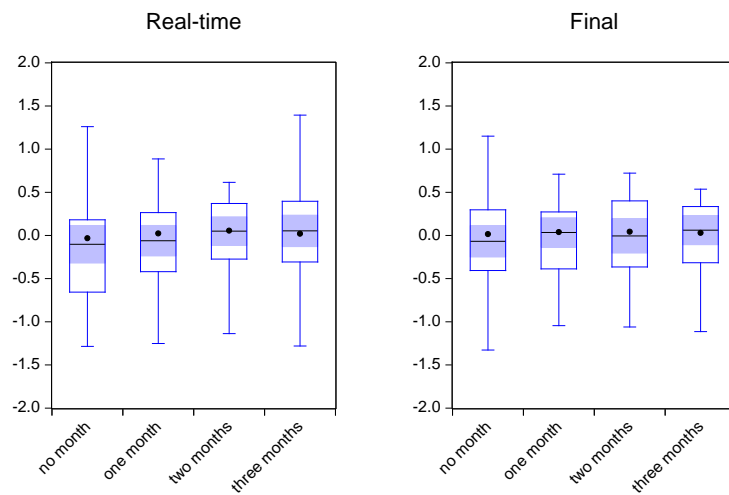
Appendix

Figure 6: Forecast errors

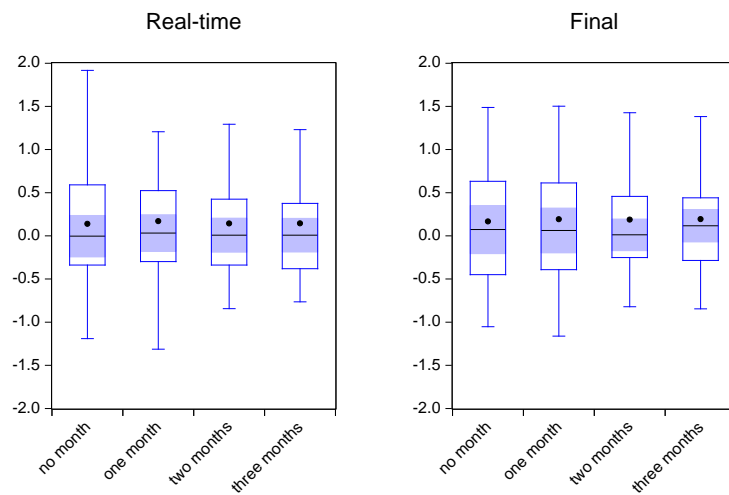
(a) Industrial production



(b) New orders

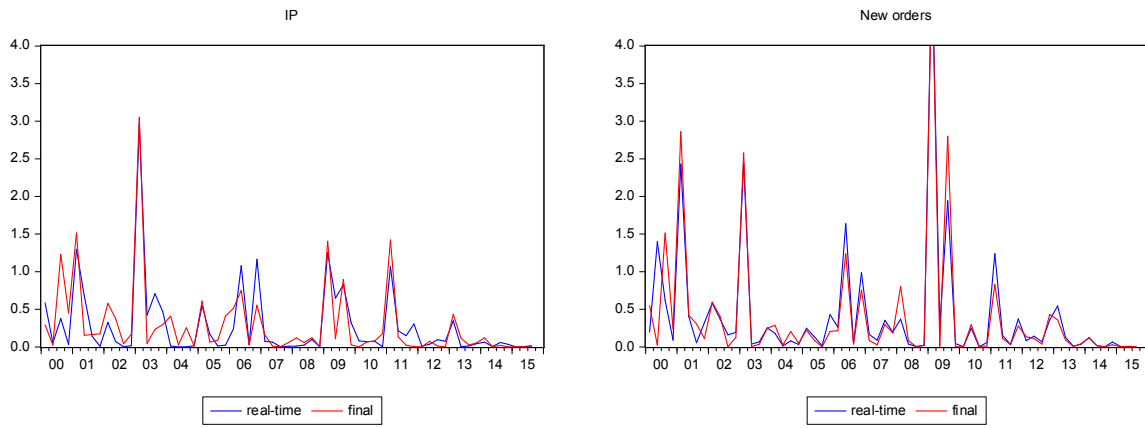


(c) Ifo situation



Note: Forecast errors based on real-time (left) and final data (right) compared with the final release are shown according to the month of availability of various indicators.

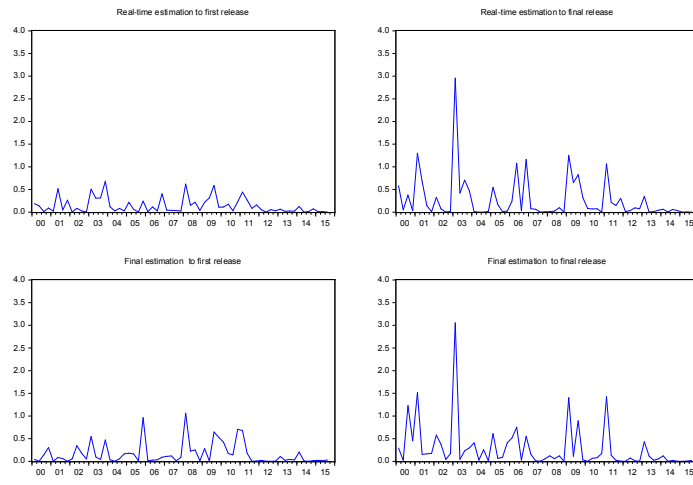
Figure 7: Forecast errors over time



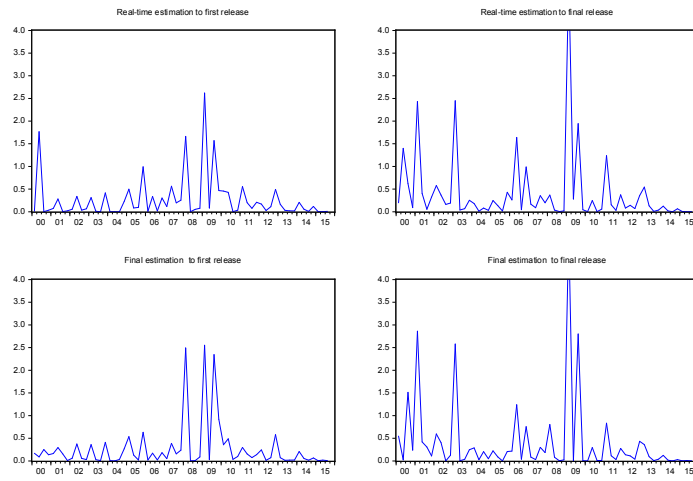
Note: Squared forecast errors based on real-time and final data compared with final-available releases are shown over time for the case where three months of information for the corresponding indicator is given.

Figure 8: Forecast errors in time

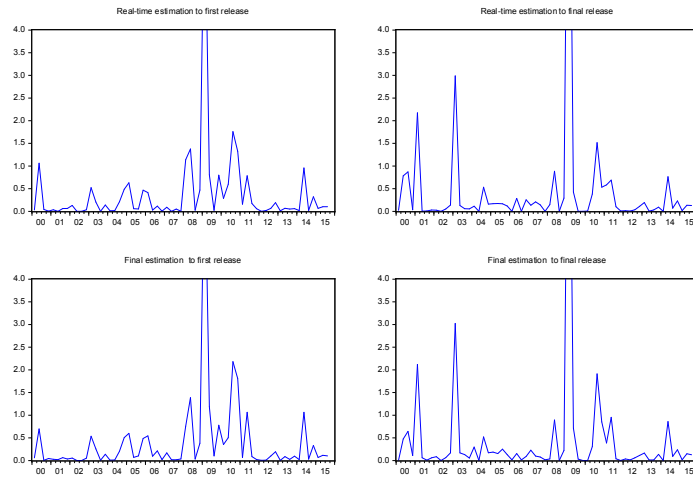
(a) Industrial production



(b) New orders



(c) Ifo situation



Note: Squared forecast errors based on real-time and final data compared with either first or final-available releases are shown over time for the case where three months of information for the corresponding indicator is given.

Halle Institute for Economic Research –
Member of the Leibniz Association

Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany

Postal Adress: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

Member of the

Leibniz
Leibniz Association