

Heinisch, Katja

Conference Paper

A real-time analysis on the importance of hard and soft data for nowcasting German GDP

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Forecasting and Information Transmission, No. B01-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Heinisch, Katja (2016) : A real-time analysis on the importance of hard and soft data for nowcasting German GDP, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Forecasting and Information Transmission, No. B01-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/145864>

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.

A real-time analysis on the importance of hard and soft data for nowcasting German GDP

Katja Heinisch* and Rolf Scheufele†

March 2016

Preliminary version

Abstract

In this paper we reexamine the relative role of soft and hard data in terms of short-term GDP forecasting. We employ mixed frequency models (MF-VARS) and real-time data to investigate the relative role of survey data relative to industrial production and orders in Germany. Special emphasis is given to the real-time data flow of surveys, production and orders. Although we find evidence that the forecast characteristics based on real-time and final data releases differ, we see only little impact on the relative forecasting performance of indicator models. However, when it comes to optimally combine soft and hard data, the use of final release data may understate the relative role of survey information.

Keywords: Mixed-frequency VAR, real-time data, forecasting

JEL Classification: C53, E37

*IWH Halle, Kleine Maerkerstrasse 8, 06108 Halle (Saale), Germany, email: Katja.Heinisch@iwh-halle.de

†Swiss National Bank (SNB), Boersenstrasse 15, 8022 Zurich, Switzerland, email: Rolf.Scheufele@snb.ch

1 Motivation

It is well known that macroeconomic forecasting face additional challenges as both target and indicator variables may be subject to revisions. This implies that early available information (e.g. first GDP releases or early available predictors for GDP) may be different from later available information for a specific point in time. This has important implications for the evaluation of forecasts from different models as forecasts typically depend on past observations via parameter estimates and starting points.

In practise, macroeconomic forecasters often ignore that the data used for conducting forecasts may change from time-to-time. Common practise is to use final data - the latest available data vintage of a given variable - to specify and evaluate a model. In this case, forecast comparisons are conducted by so-called pseudo out-of-sample forecast exercises which uses latest available data recursively for model selection, model averaging and model evaluation (e.g. Stock and Watson, 2003; Banerjee, Marcellino, and Masten, 2005). In this case, the researcher simulates the real-time situation but instead of using those data which were available at the time the forecasts are made, only a rolling or recursive estimation scheme is applied. This procedure is hardly 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 realtime data sets for many countries, there is no excuse for not using real-time data."

CROUSHORE (2011)

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

The use of real-time data comes at the cost since the number of time series to be used increases substantially and that for many countries those data are not that easily available as in the US. So real-time data should be only used when there exists any recognizable knowledge gain. Previously, Swanson (1996) compared bivariate Granger causality tests for main macroeconomic variables for different vintages of data and found that the test results vary a lot when different vintages are used. This would suggest that it 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 macroeconometric model and found sizable differences in performance when results on current vintage data are compared with latest releases. Those 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). Furthermore, for real-time policy making the first release of GDP and indicators have greater impact than subsequent releases (Sestieri, 2014).

Nowadays, more and more forecast comparisons and evaluations tend to incorporate the critique and base their analysis on full real-time data sets (see 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 it is worth to use real-time vintage data to evaluate indicator models in typical nowcasting situations. The literature on this issue is relatively unspecific. On the one hand, Stark and Croushore (2002) show that it may make a difference. Also Clements and Galvão (2009) find evidence that the forecasting accuracy of indicator models (using MIDAS) generally worsens when current vintages are used. On the other hand, Schumacher and Breitung (2008) and Banbura, Giannone, and Reichlin (2011) argue that it does not really matter whether current or final vintages are used. They find that it is much more important to mimic the publication lags of the relevant indicators (ragged-edge problem). Since both papers are using dynamic factor models, the effect of data revisions may be washed out by using a composite index. We investigate the forecasting performance of typical indicator models for nowcasting German GDP. We then document the difference between final and current data releases. Our leading indicator models consist of Mixed-Frequency (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 relative predictive power of surveys (ifo and zew) relative to hard data (industrial production, orders) when data revisions are taken into account. One question we ask is whether data revisions on industrial production and orders (which are substantially revised over time) has any impact on their relative importance compared to soft data which are rarely revised.

Does it make a difference whether macroeconomic forecasters use real-time or latest available data? Not much in general. On the one hand, we see that forecast from real-time data can systematically differ from latest release data. This can be explained by the existence of benchmark revisions which might lead to mean shifts of the entire time series. On the other hand, the relative forecasting performance is less affected whether real-time data are used or not. Although there is a tendency that hard indicators perform relatively better than surveys when latest available data are used, the general ranking is only marginally affected. Generally, we find that forecasts based on real-time estimates do reject the null hypothesis of equal forecast accuracy less often as with the latest available vintage. This reflects that forecasts based real-time data are more volatile compared to forecasts based on latest available data. Moreover, we find that latest available data are harder to predict than first releases which is consistent with the news hypothesis Mankiw and Shapiro (1986). This hypothesis implies that earlier releases get revised due to newly available information and those revisions are unpredictable from past available data.

The remainder of the paper is structured as follows. We start by briefly characterizing the data set used in the analysis. Then we present our forecasting models, followed by our procedures to evaluate the forecasts. We proceed with our main results and provide overall conclusions.

2 Data Set

Our data set consists of real-time data for Germany for quarterly seasonally-adjusted GDP as well as for selected monthly indicators from the Bundesbank. This implies that the complete time series for each real-time variable changes with the specific release data. Due to the mixed-frequency nature of the data under investigation we utilize monthly indicator releases of industrial production (ip) as well as new orders and quarterly GDP figures. While data vintages for real GDP are provided by the Deutsche Bundesbank since 1972, vintages for industrial production are available since 1995, data vintages for new orders are only available since 1999. For our real-time analysis we take into account 180 monthly data vintages from January 2000 to December 2015. Data back to 1993 are used for model estimation.

Knetsch and Reimers (2009) analyze the revision process of new orders and ip (mostly in the light of benchmark revisions). They find that first releases of growth rates on average understate the later estimates. For ip Jacobs and Sturm (2005) find some evidence that the ifo index is able to predict future revisions of ip. Taken together, this implies that first estimates of ip and new orders might be not fully optimal, but may be biased and/or be inefficient. For GDP, Faust, Rogers, and Wright (2005) report that German GDP revisions tend to be relatively large compared to revisions in other industrialized countries (at least until the 1990s). They also find some predictability for GDP revisions.

To be revised: Figure 1 shows the growth rates of GDP and indicators at the respective period according to their first release and (our) final release in December 2015. The mean revisions are almost zero for quarterly GDP growth, for the monthly indicators the mean value is about 0.15 percentage points (Table 1) summarizes the mean revisions of the time series. The results are in line with previous findings by Jung (2003) reporting a mean revision for ip -0.14 and -0.12 percentage points for orders. Hence, in both cases the first release underestimate the final one. Standard deviation are also similar to Jung (2003).

Table 1: Revision Characteristics

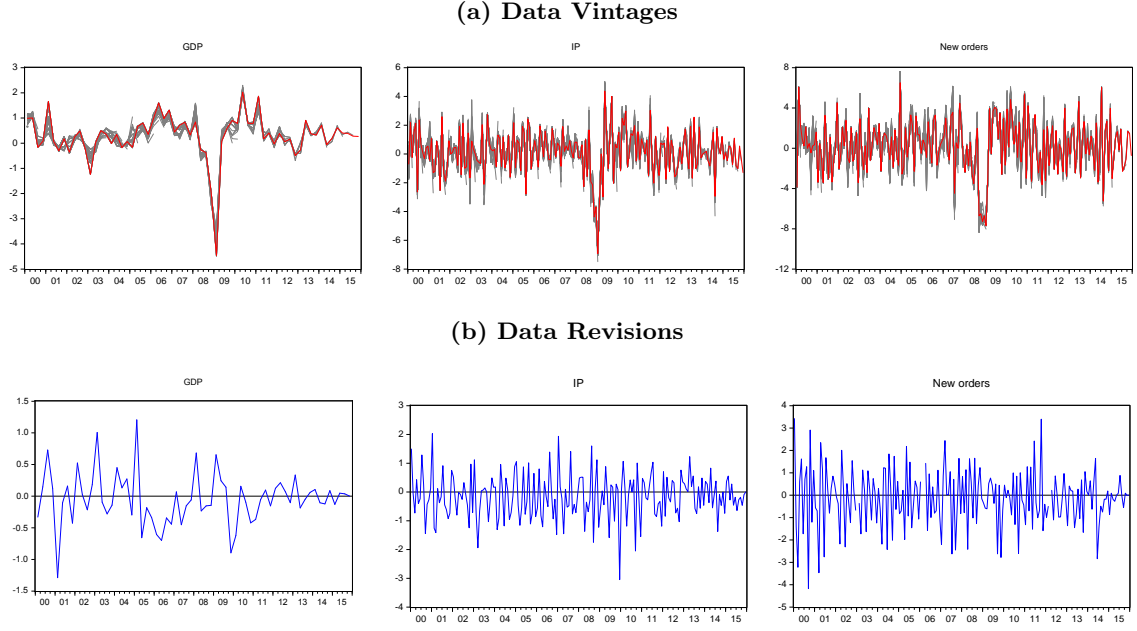
	Mean Revision	Mean Absolute Revision	Root Mean Squared Revision
GDP	-0.045 (0.051)	0.297	0.410
Industrial production	-0.150 (0.035)	0.617	0.790
New orders	-0.142 (0.055)	1.011	1.298

Note: Sample from 2000 to 2015 from the first release to the final (December 2015) data.

Beside of IP and new orders, we also take into account survey data. Namely, we employ the ifo and zew surveys. Those data are only marginally revised and we take only latest vintage data of these variables.¹

¹ In principle both series are revised as well over time, but the only major source of revision is seasonal adjustment. For the ifo indices, changes over time might also occur because of a re-weighting of sub-sectors. Clearly, employing the complete vintages would be even more realistic, we are not able to do so, because the 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 percent 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) release of the respective period. Negative values indicate that the first release was too high relative to the latest available and vice versa.

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

Another important issue when dealing with real-time data is the fact that some data are released with a certain delay. This implies that some indicators might be missing at the end of the sample (“ragged-edge”). This is particularly important when evaluating leading indicator models (Giannone, Reichlin, and Small, 2008). For instance, German GDP is published about 6 weeks after the reference quarter. Many indicators are often released more timely. On the one hand, *hard data* such as ip and new orders are released later (released with a lag of about 1.5 months) compared to *soft data* such as tendency surveys (ifo and zew) which are available within the same month. On the other hand hard data is more subject to revisions than soft data. This suggests that there is a trade-off between the two indicator groups. This issue and its consequences will be investigated below. In particular, it is the relative information content of soft versus hard data in the light of data revisions.

3 Model specifications

Leading indicator models play an important role for the assessment of the current and near term economic outlook. Typically, those 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) as well as new orders, which are observed at monthly frequency, are often considered as indicators which have some predictive power for short-term GDP predictions. Typically most of the studies do not explicitly take into account the mixed frequency nature of

the data and aggregate the higher frequent variable (the leading indicator) to quarterly frequency (Hinze, 2003; Dreger and Schumacher, 2005; Kholodilin and Siliverstovs, 2006; Drechsel and Scheufele, 2012b). However, the temporal aggregation of monthly information may be associated with a loss of information which can be sizable in typical nowcasting settings (see Wohlrabe, 2009b; Foroni and Marcellino, 2013).

To take into account the mixed-frequency problem, researchers have proposed methods that explicitly take into account different sampling frequencies (see Wohlrabe, 2009a; Kuzin, Marcellino, and Schumacher, 2011; Foroni and Marcellino, 2013, 2014).² From this discussion basically two different methods have been established that are successful in dealing with mixed-frequencies: mixed-data sampling (MIDAS) models and Mixed-Frequency (MF-)VARs. While the former method take into account higher frequent 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. The advantage of MF-VARs is that no a priori restrictions has to be imposed on the dynamics.³

In this paper we make use of MF-VARs to analyse different leading indicators for forecasting German GDP in real time. We follow Mitnik and Zadrozny (2005), Hyung and Granger (2008), Mariano and Murasawa (2010) and Kuzin, Marcellino, and Schumacher (2011) who employ a similar strategy then we do.

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

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

where x_t is 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

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

and

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

² For Germany, only a small number of papers have taken explicitly into account the mixed-frequency problem by investigating leading indicators (Mitnik and Zadrozny, 2005; Schumacher and Breitung, 2008; Wohlrabe, 2009b; Drechsel and Scheufele, 2012a).

³ See Kuzin, Marcellino, and Schumacher (2011) for pro's and cons's of MIDAS and MF-VARs and their application to forecasting euro area GDP. Recently ? provides a literature review on MF-VAR and an application to forecast US GDP.

Note that the specification of H_t implies that GDP growth x_t is only observed every third month and that it consists of month-to-month GDP growth x_t^* according 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 have to 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 this to the forecasts). Estimation of the parameters is done by maximum likelihood.⁵

After parameter estimation, we can get GDP growth predictions from the Kalman smoother, irrespectively what the real-time availability of GDP and the respective indicator is. Those forecasts are delivered iteratively.

Our estimation sample covers the period 1991M2-1999M12 and is expanding with each new monthly release of the indicators. Similarly to Schumacher and Breitung (2008) recursive estimations are conducted for the 180 vintages and hence 180 real-time forecasts (3 different forecasts for each quarter) are conducted until December 2014. To compare the real-time with pseudo real-time results, we additionally run pseudo real-time regressions, where the final vintage of GDP and

⁴ All of our results are based on a VAR(2), which was suggested by Mitnik and Zadrozny (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 As it is crucial to have good starting values for the parameter estimation in order that the estimates converge, we choose to estimate an auxiliary regression where we use 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. In the results presented below we further restrict the covariance matrix to be identical to our auxiliary regression. This helps us to achieve convergence of the likelihood in all cases. An alternative procedure would be to implement an EM-algorithm which would be computationally more demanding.

indicators, respectively, were truncated manually and the real-time situation is mimicked. Similarly to the real-time situation we also take into account the particular publication lags.

4 Forecast Evaluation

Is a model's forecasting performance different when real-time data or last releases are used to conduct forecasts? To answer this question we will proceed in two steps. First, we compare the forecasts from the two different approaches with each other. Second, we look at the relative performance of indicator forecasts relative to univariate benchmarks.

4.1 How do forecasts differ?

At an early stage we compare the model forecasts obtained from the real-time database with those of final data. Using simple measures such as plots, correlations and moments we provide a basic comparison of the two forecasting approaches. Additionally, we conduct simple tests on bias and efficiency to see whether forecasts obtained from real-time data can be seen as good predictors of forecasts obtained from the latest available vintage.

More specifically, the bias test asks whether the mean forecasts of the two approaches differ and is based on the regression:

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

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.

To see whether there is a systematic relationship between real-time forecasts and forecasts based on final data we run a Mincer-Zarnowitz type of regression, namely

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

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

4.2 Relative forecasting performance

To judge the forecast accuracy, we have to look at the forecast errors. Generally, forecast errors can be calculated in different ways. We will consider two different cases. First, we consider all the first releases of GDP as our benchmark. This benchmark takes account of the fact that in practise 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.

We will present results based on a squared loss function since we use the same criteria for model estimation. Since levels of forecast errors are hard to interpreted we follow common practise and compare the leading indicator forecasts based on the mixed frequency VARs with those of a simple univariate time series model (Granger and Newbold, 1977; Stock and Watson, 2003).⁶ This will give us an indication on how much our leading indicator model is better than the benchmark (in percentage points).

Besides the average relative performance of a model's forecast we want to ask whether a indicator based forecast is systematically better than the benchmark forecast. This has to be judged by statistical test of equal predictive ability.⁷ Given the squared loss function the corresponding null hypothesis of a indicator model i can be expressed as

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

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).

Instead of relying on parametric tests such as the DMW test, we apply nonparametric tests based on signed rank statistics. A clear advantage of the 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. This is particularly important when one analysis samples including specific events like the financial crises where loss differences are subject to unusual behaviour (large outliers and heteroscedasticity).⁸ Nevertheless, those tests have very good power characteristics compared to the DMW test (Luger, 2004).

The construction of this test statistic is straight forward with

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

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 $T \geq 20$ an approximation by normality exists. This test has been discussed and applied in similar settings by Campbell and Dufour (1995), Campbell and Ghysels

⁶ We use the AR(1) model as our benchmark univariate time-series model.

⁷ 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). Generally, tests should be valid in the case of unpredictable revisions (and non-nested models).

⁸ By looking at our loss differences ΔL_t we see very substantial deviations from normality, namely very large excess kurtosis and outliers.

(1995) and Diebold and Mariano (1995). However, one shortcoming of the basic test 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 need not be satisfied. Therefore, we follow the suggestions by Campbell and Dufour (1995) and apply a split sample techniques. We divide our original dependent loss differential into k independent ones. For instance, in the case of a MA(1) in the loss differential we would define 2 sample sets (A and B) consisting of observations $T_A = 1, 3, 5, \dots$ and $T_B = 2, 4, 6, \dots$. Then the rank-sign 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. This implies that we apply a Bonferroni correction for multiple comparisons. To decide whether the split sample technique is applied or not we use the test of serial dependence of Dufour (1981) which is also based on the rank-sign principle and shares the same robustness properties.⁹

In a second step we check whether GDP nowcasts based on hard data can be improved by combining it with other information sources. This question is related to the issue of whether combining two models result in higher forecast accuracy. Bates and Granger (1969), Clemen (1989) or 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 (Chong and Hendry, 1986). Therefore, Harvey, Leybourne, and Newbold (1998)'s test of forecast encompassing is employed.¹⁰ This test is based on the regression

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

for each indicator i . The corresponding null hypothesis equals $H_0 : \lambda = 0$, where λ is the optimal weight of combining forecast of indicator i with the ip forecast. Whenever the test rejects this implies that a combined forecast is superior relative to the single ip based model. In this case our particular interest is whether the optimal combination weights differ when real-time data are used or not.

5 Results

In this section we present our major results. First, we show how forecasts differ when models are estimated on real-time data or with final data. Second, we look at the relative performance of indicator models.

⁹ We consider serial correlation up to order 3 ($k = 3$), which translate into a maximum of 4 sub-samples. In our empirical application we find mostly evidence of $k = 0$ and $k = 1$.

¹⁰ For encompassing tests, we opted for standard parametric tests (and do not use nonparametric procedures) since there is no strong evidence of violation of the basic assumptions. Note that those tests can be formulated in terms of forecast errors instead of squared errors (which is the case in comparing forecast accuracy).

5.1 How do forecasts differ

There is generally a very high association of forecasts obtained with real-time or final data. Figure 2 presents a descriptive way to analyse the association between the two forecast procedure. The Spearman rank correlation is quite high and ranges from 0.71 (IP) to close to 1 in the case of surveys (table 2).

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

For surveys, we find a different pattern. The efficiency tests are always significant which indicates that there is no one-to-one relation between real-time and final data forecasts. Instead, GDP forecasts based on final available data are slightly more volatile. 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 have any practical implications.

5.2 Relative forecasting performance

In our analysis we can distinguish consider four different cases. The first two are concerned with the data used for estimating the models and used to conduct the forecasts. In the first case, real-time vintage data are used and in the second case final data (latest available vintage). The last two cases, concerns the benchmark used to evaluate the forecasting performance. For this case we use the latest available GDP vintage as well as first GDP releases. We start by comparing the first two cases (real-time and final data estimates) and use final GDP estimates as our benchmark.

Table 3 and figures 3 and 4 show the forecasts using real-time and final data to predict final GDP growth. Generally, we see that additional information tends to improve the forecast accuracy. This is particularly true for hard data. Most indicators show higher forecast accuracy than the univariate benchmark. The exception are the few indicators where the improvements are very small or not existent. However, in many cases the differences to the AR model are not significant.

Once we compare the difference between real-time or final data used for estimation, we see that forecast accuracy for hard data based on final data is slightly better. The overall ranking of different indicators given the specific state of information is not much affected. So we can generally say that hard data beat 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. before surveys dominate. Moreover, we see that the tests reject the null hypothesis of equal forecasting ability more often in case of final data. While for the soft data it is the other way around. ifo expectations is the only indicator that does significantly better than the benchmark in most of the cases.

Table 2: Forecast comparisons

	all months	two months	one month	no month
I. IP				
a. Bias test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = c + u_t$				
c (mean difference)	-0.133	-0.070	-0.057	-0.098
p-val	0.001	0.040	0.041	0.007
b. Efficiency test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = \alpha + \beta \hat{x}_t^{l,i} + e_t$				
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
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
II. New Orders				
a. Bias test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = c + u_t$				
c (mean difference)	-0.043	-0.009	-0.057	-0.079
p-val	0.102	0.679	0.049	0.011
b. Efficiency test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = \alpha + \beta \hat{x}_t^{l,i} + e_t$				
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
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
III. ifo situation				
a. Bias test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = c + u_t$				
c (mean difference)	-0.031	-0.028	-0.016	-0.019
p-val	0.062	0.085	0.558	0.420
b. Efficiency test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = \alpha + \beta \hat{x}_t^{l,i} + e_t$				
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
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
IV. ifo expectations				
a. Bias test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = c + u_t$				
c (mean difference)	-0.029	-0.017	-0.011	-0.006
p-val	0.072	0.311	0.352	0.684
b. Efficiency test: $\hat{x}_t^{r,i} - \hat{x}_t^{l,i} = \alpha + \beta \hat{x}_t^{l,i} + e_t$				
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
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

Note: Forecasts obtained with real-time data are compared with those obtained using final (latest available) data

When the first GDP release is the benchmark for forecasting comparisons, the picture is very similar (table 4). Worth noting is that first GDP releases are easier to predict which can be seen by the lower RMSEs of the AR model as first GDP releases are less volatile compared to later releases. This finding is consistent with the news hypothesis (Mankiw and Shapiro, 1986), which states that

Table 3: Relative RMSFE (benchmark: final GDP growth)

	M1 mid	M1 end	M2 mid	M2 end	M3 mid	M3 end	M4 mid	M4 end	M5 mid
real-time forecast to last release									
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
pseudo real-time forecast to last release									
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**. Tests on equal forecast ability are based on sign-rank tests. *, ** indicates significance at the 10% and 5% level.

first GDP releases are efficient predictors of later releases. Not surprisingly the relative errors of IP and new orders are smaller for first GDP figures as they tend to be used by the statistical office to form early GDP estimates. Therefore, forecasting gains compared to the AR can be very substantial up to 40-50% using hard data but only after a considerable time delay. Before, ifo indicators offer some improvements (in the range of 20%).

To assess whether it is worth to use other indicators and in particular survey data if we already know some IP data, we conducted encompassing tests. Table 5 indicate that orders and the ifo indicators offer significantly improvements once they are combined with IP. Generally we find that the weight is higher for surveys once the real-time data are used. Even after some months of IP is known, surveys show some value added. Generally this results implies that survey may be even more important once data revisions are taken into account.

Table 4: Relative RMSFE (benchmark: first GDP growth)

	M1 mid	M1 end	M2 mid	M2 end	M3 mid	M3 end	M4 mid	M4 end	M5 mid
real-time forecast to first release									
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
pseudo real-time forecast to first release									
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

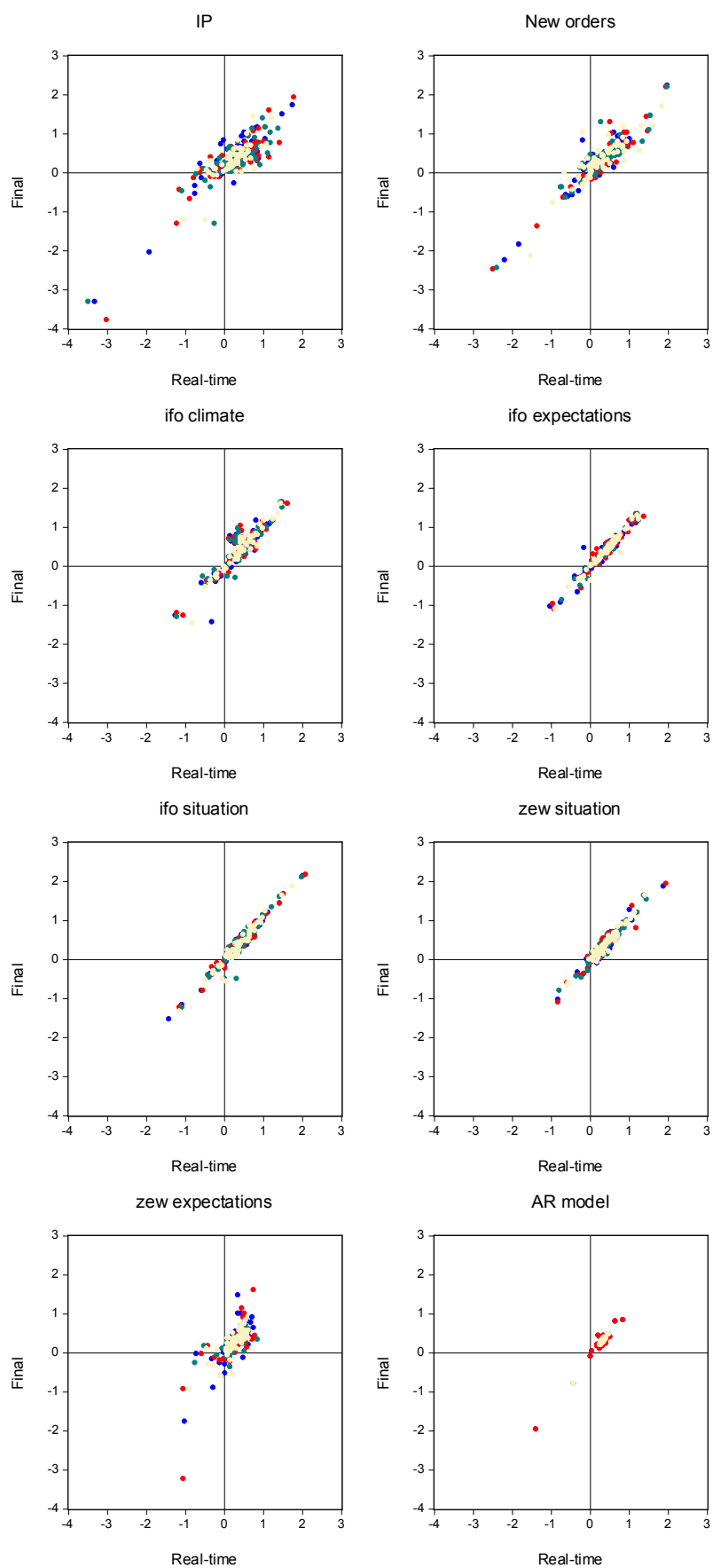
Note: Real-time and pseudo real-time forecasts compared with **first GDP releases**. Tests on equal forecast ability are based on sign-rank tests. *, ** indicates significance at the 10% and 5% level.

Table 5: Encompassing tests

	M2 mid	M2 end	M3 mid	M3 end	M4 mid
real-time forecast to last release					
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
pseudo real-time forecast to last release					
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

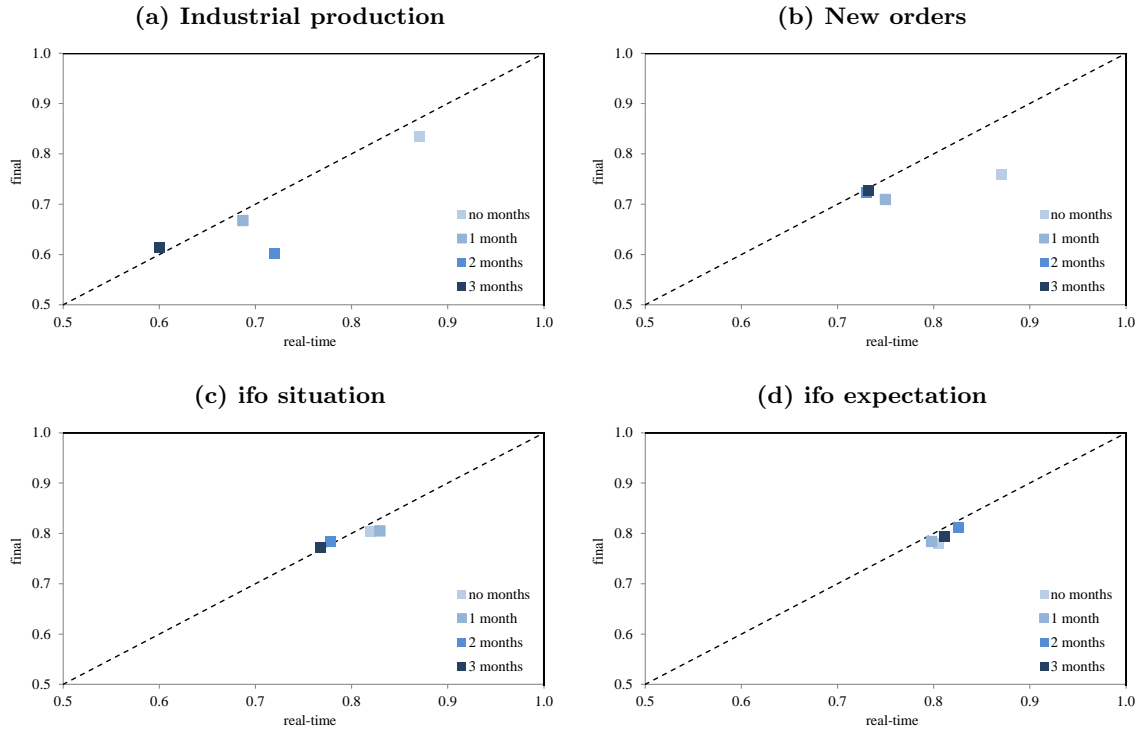
Note: The test-statistic indicates whether IP already encompasses all information or whether additional information is given by the other indicators. *** indicates significance at the 1% level.

Figure 2: Forecast comparison



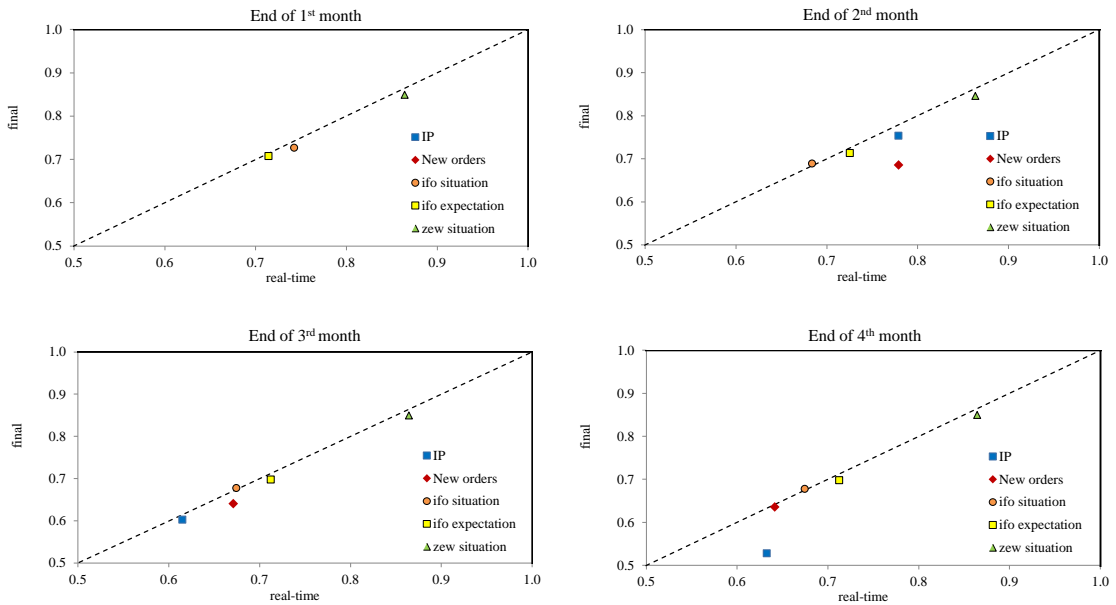
Note: Forecasts based on real-time and final data using different states of information: blue (no month), red (1 month), green (2 months) and yellow (all 3 months).

Figure 3: Forecast performance with different states of information



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

Figure 4: Forecast performance of different indicators



Note: End-of-month forecasts based on real-time and final data compared to final available releases.

6 Robustness

To check the robustness of our results we consider two additional checks. First, we investigate how your results change when forecasts are conducted using only first GDP releases. 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 it is reasonable to use first available estimates for the left-hand side when estimating the forecasting equation – even one is interested in forecasting the final (revised) data. Using a real-time-vintage estimation avoids the problem that the values of a series corresponds to different revision stages. Therefore they suggest to include only at the right-hand side of the equation data that has been available to the forecaster at this point in time. They argue that the initial release is an efficient estimate of subsequent releases, meaning that revisions are unpredictable using data at the time of the release. Table 6 shows that the relative RMSFE are marginally smaller.

Table 6: RMSFE using first GDP releases for estimation

	M1 mid	M1 end	M2 mid	M2 end	M3 mid	M3 end	M4 mid	M4 end	M5 mid
	forecast based on first releases to first release								
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 RMSFE 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). RMSFE are calculated based on first GDP releases.

The results in table 6 indicates that using only first GDP releases as the dependent variable hardly changes the conclusions relative to table 4, where models are estimated on real-time vintages. All our results show that different variants of the dependent variable have little impact on the relative forecasting performance.

6.2 Stability over time

It is well known that the financial crisis had a large impact on forecast error comparisons (Drechsel and Scheufele, 2012c). Therefore, instead of looking only at the average performance over time, we take into account potential instabilities of forecasting performance. More precisely, we follow Giacomini and Rossi (2010) and apply their fluctuation test. However, as this test is based on rolling regression of Diebold-Mariano test and as stated earlier this test might not be very reliable since our loss-differential is far from normality, we do not over-interpret this test. Since the test

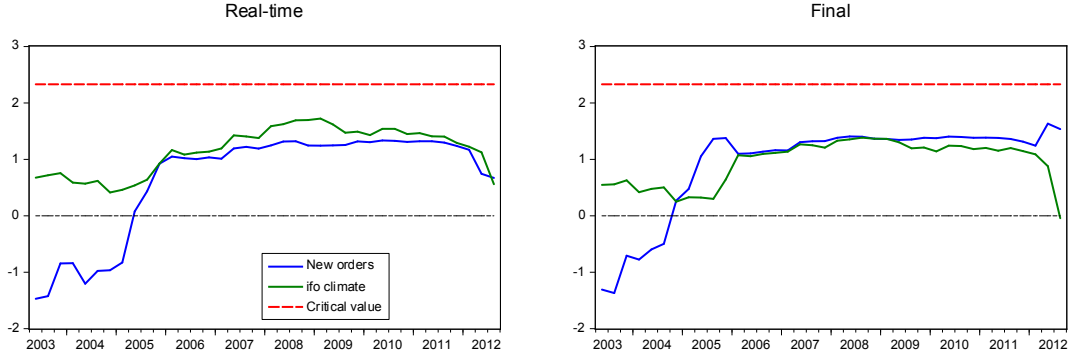
statistic is basically a weighted mean difference, we can look at the relative performance of the indicator based forecast over time.

Figure 5 shows the test statistic for different information stages (a-c) and different indicator combinations. Positive (negative) values of such differences indicate that the indicator model produces better (worse) forecasts than the AR. Intuitively we see that hard data are not much informative at the beginning of our sample. All three panels show that indicator forecasts based on hard data are inferior to AR forecasts. However, since 2005/2006 the relative forecasting performance using hard data has improved. Most recently the forecasting performance of survey data has deteriorated. Generally, the forecasting performance over time using real-time data or final data is very similar.

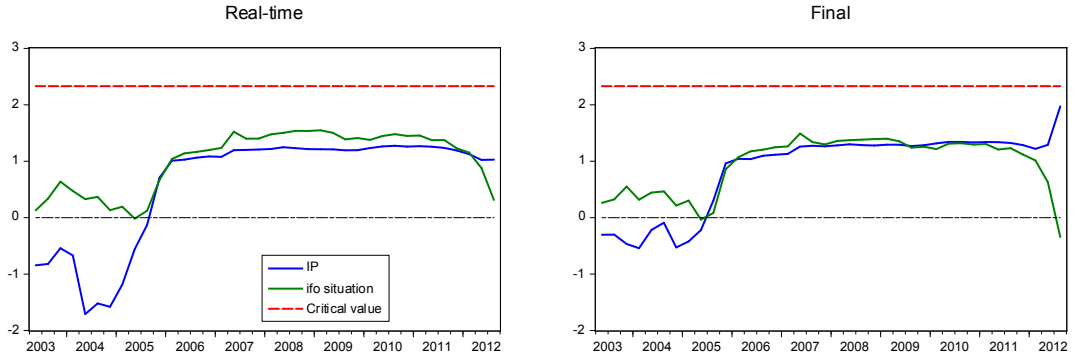
[...]

Figure 5: Local relative performance based on fluctuation tests

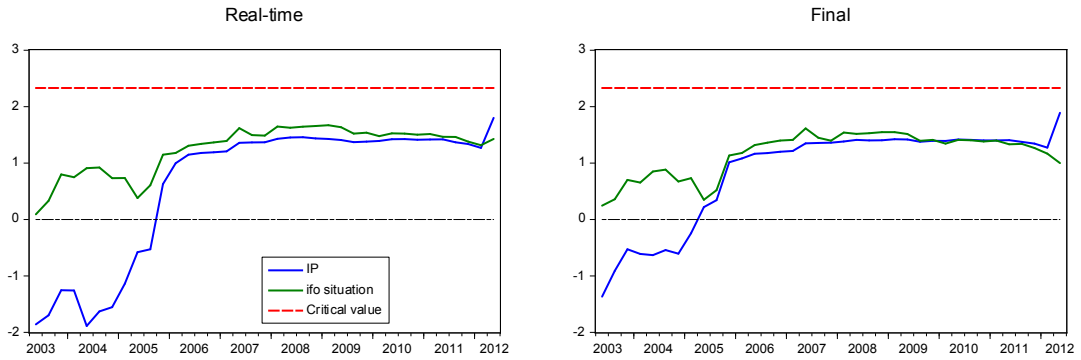
(a) end of 2nd month



(b) end of 3rd month



(c) end of 4th month



Note: Local relative performance as the out-of-sample MSFE differences 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

[...]

References

- 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.
- 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, 663–694.
- CAMPBELL, B., AND J.-M. DUFOUR (1995): “Exact Nonparametric Orthogonality and Random Walk Tests,” *The Review of Economics and Statistics*, 77, 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, 17–31.
- CHONG, Y. Y., AND D. F. HENDRY (1986): “Econometric Evaluation of Linear Macro-Economic Models,” *The Review of Economic Studies*, 53, 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, 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, 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): “Bottom-up or Direct? Forecasting German GDP in a Data-rich Environment,” Working Papers 2012-16, Swiss National Bank.

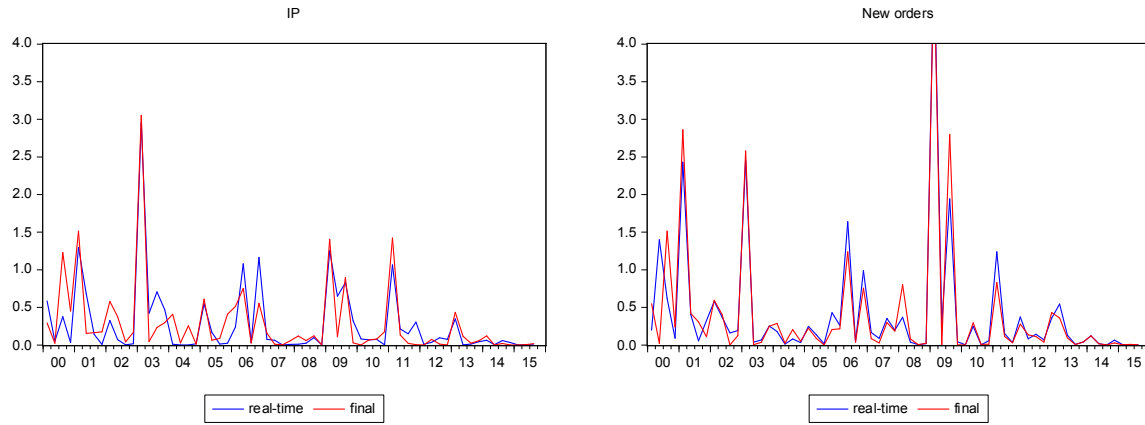
- (2012b): “The Financial Crisis from a Forecaster’s Perspective,” *Kredit und Kapital*, 2012(1), 1–26.
- (2012c): “The Performance of Short-term Forecasts of the German Economy before and during the 2008/2009 Recession,” *International Journal of Forecasting*, 28(2), 428 – 445.
- DREGER, C., AND C. 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, 403–419.
- 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, 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.
- 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, 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. SILVERSTOVS (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, 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,” Discussion paper.
- MANKIW, N., AND M. SHAPIRO (1986): “News or noise: an analysis of GDP revisions,” *Survey of Current Business*, 66, 20–25.
- MARIANO, R., AND Y. MURASAWA (2010): “A Coincident Index, Common Factors, and Monthly Real GDP,” *Oxford Bulletin of Economics and Statistics*, 72, 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*, forthcoming(just-accepted).
- 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, 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, 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

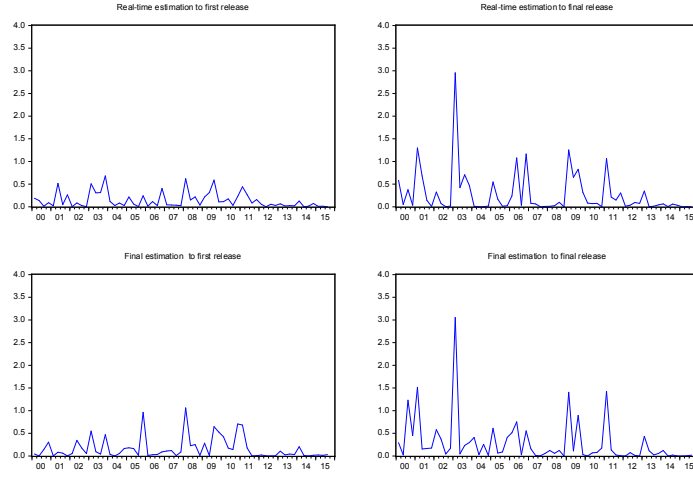
Figure 6: Forecast Errors over Time



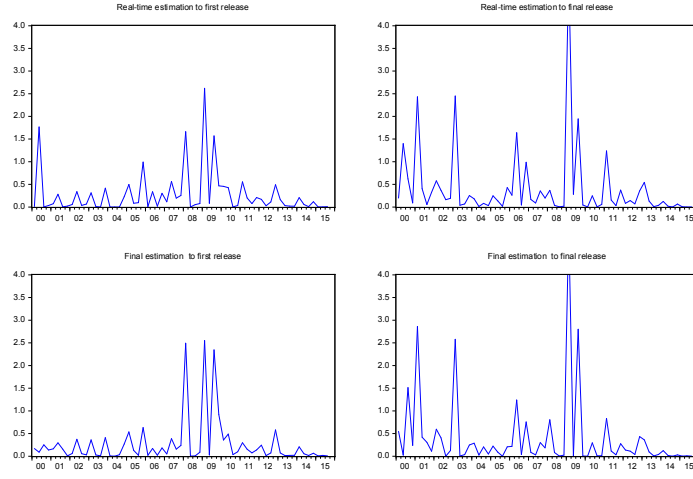
Note: Squared forecast errors based on real-time and final data compared to final available releases are shown over time for the case that three month of information for the corresponding indicator is given.

Figure 7: Forecast Errors in Time

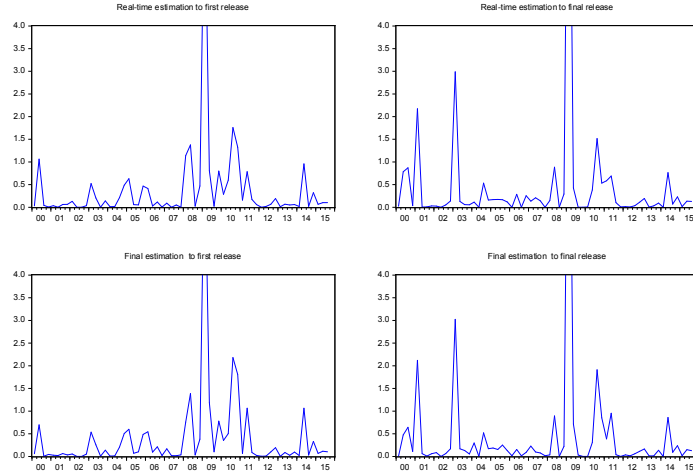
(a) IP



(b) New orders



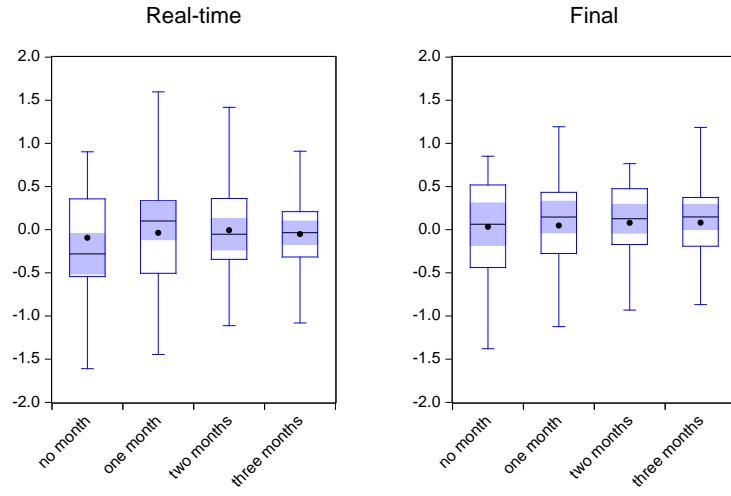
(c) ifo situation



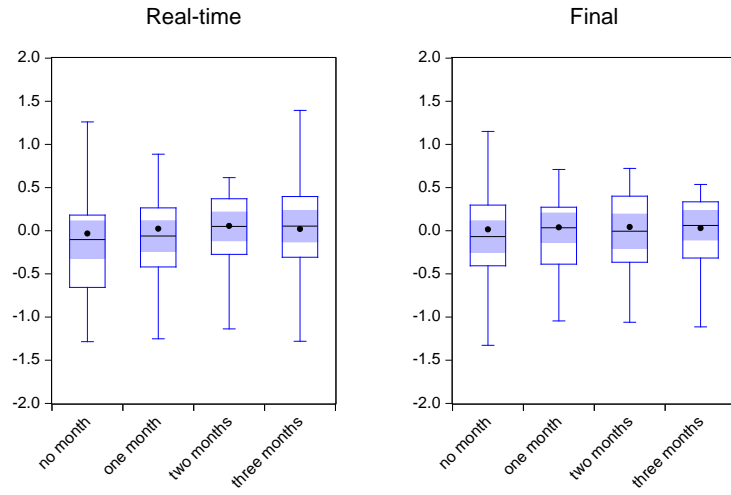
Note: Squared forecast errors based on real-time and final data compared either to first or final available releases are shown over time for the case that three month of information for the corresponding indicator is given.

Figure 8: Forecast Errors

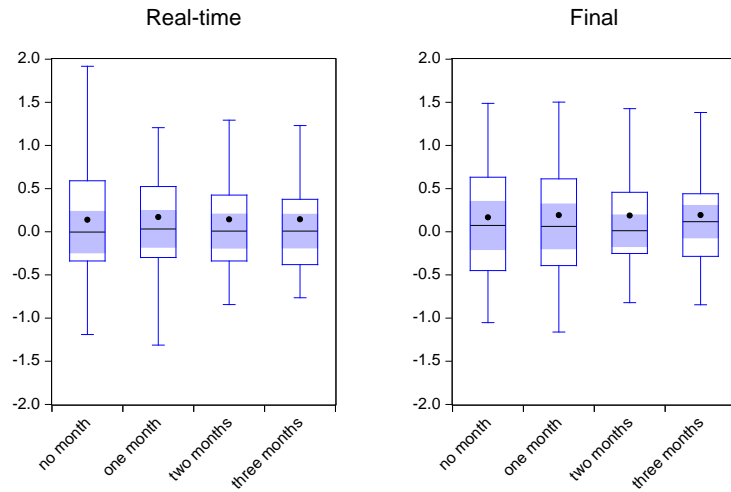
(a) IP



(b) New orders



(c) ifo situation



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