Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 24 (2008) 386–398
www.elsevier.com/locate/ijforecast

Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data

Christian Schumacher [a,1], Jörg Breitung [b,*]

[a] Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main, Germany
[b] Institute of Econometrics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany

Abstract

This paper discusses a factor model for short-term forecasting of GDP growth using a large number of monthly and quarterly time series in real-time. To take into account the different periodicities of the data and missing observations at the end of the sample, the factors are estimated by applying an EM algorithm, combined with a principal components estimator. We discuss some in-sample properties of the estimator in a real-time environment and propose alternative methods for forecasting quarterly GDP with monthly factors. In the empirical application, we use a novel real-time dataset for the German economy. Employing a recursive forecast experiment, we evaluate the forecast accuracy of the factor model with respect to German GDP. Furthermore, we investigate the role of revisions in forecast accuracy and assess the contribution of timely monthly observations to the forecast performance. Finally, we compare the performance of the mixed-frequency model with that of a factor model, based on time-aggregated quarterly data.

JEL classification: E37; C53

Keywords: Forecasting; GDP; EM algorithm; Principal components; Factor models

1. Introduction

Macroeconomic policy-making in real-time faces the problem of needing to assess the current state of the economy with incomplete statistical information. Important economic variables are released with considerable time lags. As a key indicator of real economic activity, GDP is published at a quarterly frequency, and only with a considerable delay. For example, in Germany, an initial GDP estimate is released about 6 weeks after the end of the respective quarter. If policy decisions require information within or right after that quarter, more timely information is needed. Such timely information might be available as observations of monthly and quarterly variables, and could help to either forecast current quarter GDP, the so-called nowcast, or make forecasts at longer horizons.

* Corresponding author.
E-mail addresses: christian.schumacher@bundesbank.de (C. Schumacher), breitung@uni-bonn.de (J. Breitung).
[1] Tel.: +49 9566 2939, +69 9566 2939; fax: +49 9566 2982, +69 9566 2982.

0169-2070/$ - see front matter © 2008 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2008.03.008

In general, real-time forecasting tends to be characterized by three features: Firstly, due to revisions, the data actually available in a particular month may differ substantially from the final values released later by statistical offices. Thus, using the vintages of data that were actually available to forecasters in the past is desirable for forecast comparisons. Secondly, the dataset may suffer from the ragged-edge problem in econometrics, namely be missing values of some of the variables at the end of the sample due to publication lags. Thirdly, some of the indicators are available only at a quarterly frequency, whereas others are available at a monthly or even higher frequency. Thus, a mixed-frequency problem also has to be resolved.

In this paper, we address these real-time issues with regard to forecasting German GDP growth in a factor model framework. Following Stock and Watson (2002a), we estimate the factors by using principal components. To take into account data with different frequencies and with missing values at the end of the sample due to publication lags, an expectation-maximisation (EM) algorithm is employed. The EM algorithm can interpolate missing observations at the end of the sample, as far as timely information from other high-frequency indicators is available. However, for long forecast horizons, additional forecasting methods are necessary. We discuss alternative projection methods and apply them in our empirical application. The dataset in the empirical application is a novel real-time dataset for post-unification Germany, consisting of about 50 quarterly and monthly time series. Time series subject to statistical data revisions are available as monthly vintages.

In the empirical application, the vintages of data are used to recursively forecast German GDP. In order to assess the importance of data revisions, we compare real-time forecasts with forecasts from using final-vintage data. In addition, the impact of timely monthly observations on forecast accuracy is investigated by appropriately modifying the real-time dataset. Furthermore, we compare the mixed-frequency factor model with a quarterly single-frequency factor model and other simple benchmark models to see whether the more detailed mixed-frequency information can be exploited successfully.

In the literature, mixed-frequency forecasting is often carried out in a state-space framework, see Mariano and Murasawa (2003), Evans (2005), Mittnik and Zadrozny (2005) and Nunes (2005), where the high-frequency model and the frequency conversion can be addressed in one coherent framework. The interpolation approach, such as that of Friedman (1962), Chow and Lin (1971) and Mitchell, Smith, Weale, Wright, and Salazar (2005), allows for inter- and extrapolation in a high-frequency regression model with mixed-frequency data. Other approaches rely on bridge equations, e.g. Baffigi, Golinelli, and Parigi (2004), where a dynamic equation is estimated between the low-frequency target variable and time-aggregated indicators. Separate high-frequency time series models are applied to provide forecasts of the high-frequency indicators, and the forecast values are time-aggregated and plugged into the bridge equation to obtain the forecast of the low-frequency variable. Another line of research is based on mixed data-frequency sampling, as proposed by Ghysels, Sinko, and Valkanov (2007) and Clements and Galvão (2007), where a low-frequency variable is explained by high-frequency indicators using distributed lag regression models. All of these approaches are typically restricted to a rather limited information set with a small number of variables, whereas forecasts based on principal components as factors allow for very large datasets. Giannone, Reichlin, and Small (2005) employ a large monthly factor model, with special emphasis on the information content of different groups of monthly indicators for GDP. An alternative approach based on factor models is proposed by Altissimo, Cristadoro, Forni, Lippi, and Veronese (2006), where dynamic factors estimated from a large number of monthly indicators are used to forecast low-frequency fluctuations in GDP growth. Our application is more closely related to that of Bernanke and Boivin (2003), who investigate the real-time forecasting accuracy of the factor model proposed by Stock and Watson (2002a) for US data. However, their comparison of alternative forecast procedures is concerned with forecasting monthly variables, whereas our focus is on forecasting the quarterly variable GDP. The interpolation and backcasting properties of the EM algorithm are discussed by Angelini, Henry, and Marcellino (2006), and a comparison to other approaches is provided by Marcellino (2007).

The paper is organized as follows. In Section 2 we introduce the EM algorithm for the estimation of factors based on mixed-frequency data, and outline alternative forecasting methods to be used in this context. Section 3 presents the real-time dataset and the design of the forecast experiment, and discusses the results of the empirical application in detail. Section 4 concludes.

2. Factor forecasting GDP with mixed-frequency data

This section outlines the estimation and forecasting procedure based on a factor model, where two cases are distinguished: the case without data irregularities (Section 2.1) and situations with mixed-frequency data and missing observations using the EM algorithm (Section 2.2). Furthermore, alternative forecasting techniques based on the estimated factors are also proposed in Section 2.3.

2.1. Factor model for single-frequency data

Let $x_{i,t}$ denote the ith variable from a collection of stationary time series $\{x_{i,t}\}_{i=1,\ldots,N}$. The index t indicates equidistant time intervals, so different periodicities are neglected so far. We assume that the variables in the dataset can be represented as the sum of two mutually orthogonal components: the common and idiosyncratic components. The common component of each variable is a linear combination of a small number of common factors collected in the $r\times 1$ vector F_t. The idiosyncratic components $e_{i,t}$ are variable-specific. Thus, each variable can be represented as

$$x_{i,t} = \Lambda_i' F_t + e_{i,t} \tag{1}$$

for $t=1,\ldots,T$ and $i=1,\ldots,N$. The $r\times 1$ vector Λ_i collects the factor loadings of the r factors. For later use in the EM algorithm, the T time observations of the ith variable are represented in matrix notation as

$$X_i = F\Lambda_i + e_i, \tag{2}$$

where the $T\times 1$ vectors $X_i=(x_{i,1},\ldots, x_{i,T})'$ and $e_i=(e_{i,1},\ldots, e_{i,T})'$, and the $T\times r$ matrix $F=(F_1,\ldots, F_T)'$, stack the observations from all time periods. It is assumed that F_t is generated by the stationary VAR(p) model

$$F_t = \Phi_1 F_{t-1} + \cdots + \Phi_p F_{t-p} + u_t, \tag{3}$$

where $u_t\sim N(0,\ \Omega)$, and all roots of the autoregressive polynomial $\Phi(L)=I-\Phi_1 L-\cdots-\Phi_p L^p$ are outside the unit circle. The idiosyncratic components are mutually and serially uncorrelated random variables, with $e_t=(e_{1,t},\ldots,\ e_{N,t})'\sim N(0,\ \Sigma)$, where $\Sigma=\text{diag}(\sigma_1^2,\ldots,\ \sigma_N^2)$. Furthermore, the common factors and idiosyncratic components are assumed to be orthogonal, i.e. $E(u_t e_{i,t-k})=0$ for all i and k.

Following Chamberlain and Rothschild (1983), the assumption of a strict factor model can be relaxed by allowing for some non-pervasive cross-sectional correlation of the idiosyncratic components, which leads to the so-called approximate factor model. Furthermore, Stock and Watson (2002a,b) and Bai and Ng (2002) show that even under weak serial correlation and heteroscedasticity of the idiosyncratic components and additional regularity conditions, the factors and factor loadings can be estimated consistently. Nevertheless, for the subsequent analysis we will confine ourself to the strict factor model, as this framework allows the estimation of missing values using the EM algorithm.

The factors in this model can be estimated consistently by the method of principal components (PC), see Stock and Watson (2002b) and Bai and Ng (2002). To eliminate scale effects, the time series are standardized to have zero mean and unit variance in-sample prior to the PC decomposition. Let V denote the $N\times r$ matrix of the r eigenvectors corresponding to the r largest eigenvalues of the sample correlation matrix. The PC estimators of the factors and the factor loadings are obtained as $\hat{F} = XV/\sqrt{N}$ and $\hat{\Lambda} = V\sqrt{N}$, where $X=(X_1,\ldots, X_N)$ and $\hat{\Lambda}=(\hat{\Lambda}_1,\ldots, \hat{\Lambda}_N)'$.

2.2. The EM algorithm for unbalanced data

As was mentioned in the introduction, forecasting in real-time implies that the dataset may be highly unbalanced due to missing observations and different sampling frequencies. The EM algorithm, together with PC decomposition, as introduced by Stock and Watson (2002a), can tackle such data irregularities in a factor model framework. Furthermore, by virtue of recursive application of the EM algorithm to real-time vintages of data, it can also handle revisions indirectly.

2.2.1. Mapping from unobserved to observed GDP growth

In Section 2.1, the dataset was implicitly assumed to be balanced, without any data irregularities. Now, assume that some elements of the $T\times 1$ vector X_i are unobserved or missing. Let X_i^{obs} denote the $T^{\text{obs}}\times 1$ vector of available observations for the ith variable. A key step during the EM iterations is the mapping from the (partially) unobserved monthly values X_i to the observed values X_i^{obs}. This mapping can be written as a linear relationship

$$X_i^{\text{obs}} = A_i X_i, \tag{4}$$

where A_i is a known $T^{\text{obs}} \times T$ aggregator or selection matrix that can tackle missing values or different frequencies depending on the particular variable, see Stock and Watson (2002a, pp. 156–157). For the steps of the EM algorithm described below, a similar mapping is needed for all variables subject to data irregularities, and we require a full row rank of A_i in each case. Without data irregularities, A_i is simply the identity matrix.

Concerning the main variable of interest, GDP, we follow Mariano and Murasawa (2003) and estimate unobserved month-on-month GDP growth. According to Eq. (4), we require a mapping from unobserved monthly GDP growth to observed quarterly GDP growth. To this end, we approximate quarterly GDP growth by the first difference of logged GDP, and denote this variable as y_t^q. Since GDP is a flow variable, y_t^q is related to the the unobserved monthly growth rate y_t^m by

$$y_t^q = \frac{1}{3}\left(y_t^m + 2y_{t-1}^m + 3y_{t-2}^m + 2y_{t-3}^m + y_{t-4}^m\right). \tag{5}$$

The time index t denotes a calendar month, and Eq. (5) holds for $t=3, 6, 9, \ldots, T$, assuming for simplicity that new quarterly observations are available in the last month of the quarter. We stack the time series observations, yielding, the relationship $Y^q = A_y Y^m$ in line with Eq. (4), where

$$A_y = \frac{1}{3}\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 \\ \cdots & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 \end{pmatrix}, \tag{6}$$

$Y^q = (\ldots, y_{T-6}^q, y_{T-3}^q, y_T^q)'$, and the monthly growth rates $Y^m = (\ldots, y_{T-2}^m, y_{T-1}^m, y_T^m)'$.

Missing values at the end of the sample can also be considered in this framework. In the case of missing quarterly observations, the rows of matrix A_y corresponding to missing values of the quarterly GDP time series have to be removed. Assume, for example, that the final growth rate of GDP is missing at the end of the sample, that is $Y^q = (\ldots, y_{T-6}^q, y_{T-3}^q)$, whereas it is assumed that monthly indicators are available up to the final month T. In this case, the bottom row of Eq. (6) has to be removed.

2.2.2. Steps of the EM algorithm

In the following, the steps of the estimation procedure are described in more detail.

1. Provide initial values of the missing observations and initial monthly values of the quarterly data.[2] These values, together with the monthly time series not subject to data irregularities, comprise a balanced monthly dataset denoted by $\hat{X}^{(0)}$, a $T \times N$ matrix. We estimate monthly factors $\hat{F}^{(0)}$ and loadings $\hat{\Lambda}^{(0)}$ as in the single-frequency case in Section 2.1.
2. *E-step:* For the jth iteration and the ith variable, compute updated values in X_i by the expectation of X_i conditional on the observed values X_i^{obs}, factors $\hat{F}^{(j-1)}$ and loadings $\hat{\Lambda}_i^{(j-1)}$ from the previous iteration $j-1$ (or estimates of the initial step 1) using

$$\begin{aligned}\hat{X}_i^{(j)} &= \hat{F}^{(j-1)}\hat{\Lambda}_i^{(j-1)} \\ &\quad + A_i'(A_i A_i')^{-1}\left(X_i^{\text{obs}} - A_i \hat{F}^{(j-1)}\hat{\Lambda}_i^{(j-1)}\right),\end{aligned} \tag{7}$$

   see Stock and Watson (2002a, p. 156). Matrix A_i is the variable-specific selection or aggregator matrix from Eq. (4).
3. Repeat step 2 for all variables in the sample that contain missing values or have to be estimated, and collect all estimated values and observations in the matrix $\hat{X}^{(j)}$.
4. *M-step:* The monthly data $\hat{X}^{(j)}$ are used to obtain new PC estimates of the factors $\hat{F}^{(j)}$ and loadings $\hat{\Lambda}^{(j)}$. The factors and loadings are returned to step 2 above, and the subsequent steps are repeated until

[2] To obtain initial values, we follow Biernacki, Celeux, and Govaert (2003), and employ random initialisation with previous EM runs based on a fixed small number of iterations. For all time series with data irregularities, we replace missing observations by random draws from the standard normal distribution. An EM algorithm run with five iterations provides an estimate of the average idiosyncratic variance. We repeat this ten times, and choose the data corresponding to the smallest idiosyncratic variance to be the initial data. Note that the data is standardised to have zero mean and unit variance prior to the PC decomposition.

convergence. In our application below, the algorithm stops if the maximum percentage change of the variables' estimates is smaller than 10^{-5}.

We denote the final monthly factor estimates by $\hat{F}_t=\hat{F}^{(J)}$ for $t=1,\ldots, T$, and the final loadings as $\hat{\Lambda}=\hat{\Lambda}^{(J)}$, where J is the final iteration of the EM algorithm after convergence. Similarly, the monthly final-iteration data values are $\hat{X}_t=\hat{X}_t^{(J)}$. The estimated monthly values of GDP, $\hat{y}_t^m$, can be taken directly from the particular element of $\hat{X}_t$, and $\hat{\Lambda}_y$ is the corresponding GDP loading vector from $\hat{\Lambda}$.

2.3. Forecasting monthly GDP

The EM algorithm provides monthly values for GDP for those quarters where quarterly figures of GDP are available. Furthermore, if the GDP release is not yet available, but some monthly indicators are, the EM algorithm can be employed to obtain an estimate of the corresponding monthly GDP. Thus, the EM algorithm already provides a forecast, although it might also be labelled an 'extrapolation' following Chow and Lin (1971), for example. In practice, however, monthly indicators are only available up to a few months in advance of the GDP. For longer forecast horizons, additional forecasting techniques have to be invoked. Although factor-based forecasting using single-frequency data is well known in the literature (see the comprehensive discussion in Boivin & Ng, 2005), we have to modify the existing methods for forecasting quarterly GDP with monthly factors in the mixed-frequency framework chosen here. Assume that we are interested in the conditional forecast of monthly GDP growth in month $T+h$, denoted by $y^m_{T+h|T}$, where T denotes the latest time period where a monthly indicator is available. To be comparable to quarterly GDP observations later, the forecasts of monthly GDP are aggregated to quarterly figures by using Eq. (5).

According to Eq. (1), the theoretical factor forecast $y^m_{T+h|T}$ is obtained as

$$y^m_{T+h|T} = \Lambda_y' F_{T+h|T}, \tag{8}$$

where we make use of the fact that the idiosyncratic component is assumed to be white noise. The dynamic model of the factors in Eq. (3) can help to express the forecast in terms of in-sample observations of the factors according to the Wiener–Kolmogorov prediction formula

$$F_{T+h|T} = B^h(L)F_T, \tag{9}$$

where

$$\begin{aligned} B^h(L) &= B_0^h + B_1^h L + \cdots + B_{p-1}^h L^{p-1} \\ &= \left[\Phi(L)^{-1}/L^h\right]_+ \Phi(L), \end{aligned}$$

and $[\cdot]_+$ denotes the annihilation operator that sets negative powers of L equal to zero.[3] Note that the matrices B_j^h are nonlinear functions of the matrices Φ_1, …, Φ_p for $j=0,\ldots, p^{-1}$.

For empirical applications, alternative approaches can be used to forecast the factors. In the single-frequency framework, Boivin and Ng (2005) distinguish three different approaches. We extend their approaches to the mixed-frequency case in the following way:

1. The iterative multi-step forecast (IMS) is based on the estimated VAR polynomial $\hat{\Phi}(L)$, yielding

$$\begin{aligned} \hat{F}^{\text{IMS}}_{T+h|T} &= \hat{B}^h(L)\hat{F}_T \\ &= \left[\hat{\Phi}(L)^{-1}/L^h\right]_+ \hat{\Phi}(L)\hat{F}_T. \end{aligned} \tag{10}$$

The forecast of monthly GDP is obtained as $\hat{y}^m_{T+h|T}=\hat{\Lambda}_y'\hat{F}^{\text{IMS}}_{T+h|T}$ according to the factor representation in Eq. (8).

2. The direct multi-step (DMS) forecast (see also Chevillon & Hendry, 2005) is obtained from the direct and linear multivariate regression

$$\hat{F}_{t+h} = C^h(L)\hat{F}_t + v^h_{t+h}, \tag{11}$$

with $C^h(L)=C_0^h+C_1^hL+\cdots+C^h_{p-1}L^{p-1}$ for $t=p, p+1$, …, $T-h$. Let $\hat{C}^h(L)$ denote the least-squares estimator of $C^h(L)$. The DMS forecast is obtained as

$$\hat{F}^{\text{DMS}}_{T+h|T} = \hat{C}^h(L)\hat{F}_T, \tag{12}$$

and the monthly GDP forecasts are $\hat{y}^m_{T+h|T}=\hat{\Lambda}_y'\hat{F}^{\text{DMS}}_{T+h|T}$. In large samples, $\hat{C}_j^h$ converges in probability to B_j^h, and therefore the IMS and DMS approaches yield similar forecasts. However, in small samples, the DMS

[3] See Sargent (1987, p. 328).

approach is inefficient relative to the IMS approach, since the error v_t^h in Eq. (11) is autocorrelated for $h>1$.

3. A forecast that uses only the estimated factors and neglects the other assumptions underlying the factor model is the so-called unrestricted (U) forecast, so labelled by Boivin and Ng (2005). In the context of monthly GDP forecasting, we have

$$\hat{y}^m_{T+h|T} = \hat{D}^h(L)\hat{F}_T. \tag{13}$$

The estimate of the lag polynomial $\hat{D}^h(L)$ is provided by regressing $\hat{y}_t^m$ onto $\hat{F}_{t-h}$ and lag estimates $\Lambda_y' \; B^h(L)$. Thus, the U forecast can be seen as a DMS forecast that does not rely on the estimate of the loading matrix $\hat{\Lambda}_y'$, in contrast to the IMS and DMS forecasts.

As shown by Boivin and Ng (2005), the IMS, DMS and U forecasts may perform differently. Marcellino, Stock, and Watson (2006) and Chevillon and Hendry (2005) provide a thorough comparison of the relative advantages of iterative and direct forecasting. Below, we follow the literature on factor-based forecasting, and compare the alternative methods above in the present mixed-frequency application.

3. Real-time forecasting of German GDP

Having discussed different forecasting alternatives for GDP, we now provide an empirical application using German real-time data.

3.1. German real-time data

We use a composite real-time dataset of post-unified Germany. We predict German GDP growth for the time period 1999Q2 to 2005Q1, providing us with 24 observations for forecast comparison. The first period with GDP data available is 1991Q2. In addition to GDP, the dataset includes quarterly demand components of GDP, gross value added, industrial production and incoming orders by sectors. We also consider a variety of financial indicators, such as interest rates, stock price indices, and exchange rates, as well as survey data on business confidence and expectations. Overall, the dataset consists of 52 time series, 39 monthly series and 13 quarterly series. The dataset is somewhat smaller than the data sets of other applications using large factor models, but is considerably larger than those used in applications of state-space or bridge-equation models.[4] Furthermore, the results obtained by Boivin and Ng (2006) show that the sample size of the dataset has only a minor effect on estimation.[5] For monthly (quarterly) data, the sample starts 1991M4 (1991Q2). The starting dates are the same for all vintages of data and correspond to the GDP data. The final dates in each vintage of data vary considerably across the different time series, and are determined by the publication delay of each series. For example, the data of vintage April 2005 contains GDP data up to the fourth quarter 2004, and hence has a release delay of 3 months compared with the calendar. Production indices and incoming orders are available up to February 2005. The financial and survey time series are available up to March 2005. Thus, a vintage in the dataset contains a highly unbalanced dataset, with missing values at the end of the sample. More detailed information about the time series can be found in Appendix A.1.

3.2. Design of the recursive real-time forecast experiment

In our forecast exercise, we forecast German GDP recursively. Every month, a new vintage of monthly and possibly quarterly data becomes available. The factor model is reestimated with the extended dataset, and forecasts are computed. When estimating the factors, the different availabilities of the data, as discussed in the data section above, are automatically taken into account by the EM algorithm, as described in Section 2. Prior to factor estimation, we apply a number of transformations to the data, see Stock and Watson (2002a, p.149). Natural logarithms were taken for all time series except interest rates, and stationarity was obtained by appropriately differencing the time series. To eliminate scale effects when estimating the factors, the time series were standardized to have zero mean and unit sample variance.

As we have monthly indicators in addition to the quarterly indicators, and the factor model allows for

[4] In their large factor model for the USA, Bernanke and Boivin (2003, pp. 529–531) use about 80 time series, whereas Nunes (2005) estimates a six-variable state-space model for monthly GDP.

[5] See also the theoretical discussion in Heaton and Solo (2006, pp. 12–14).

mixed frequencies, we can forecast quarterly GDP at monthly forecast horizons. Here, we use a monthly forecast horizon of 6 months. We also experimented with longer horizons; however, no informative forecasts were obtained for these. The forecasts for the first three months can be regarded as nowcasts of GDP (see Giannone et al., 2005, for an extensive discussion on nowcasting). In particular, horizon one is a nowcast for the current quarter of GDP based on information available in the third month of that quarter, and horizon two corresponds to a nowcast based on information from the second month. Horizons three to six are one-quarter ahead forecasts. For example, horizon four corresponds to a GDP forecast of a particular quarter, based on information available in the third month of the previous quarter.[6]

For forecasting using the factor model, both the number of factors and the lag orders of the projections have to be specified. In the empirical literature, there is considerable uncertainty about the appropriate choice of the number of factors, since information criteria seem to give misleading results in some cases. For example, Bernanke and Boivin (2003, see footnote 7) use three factors for their real-time applications for the US, whereas the Federal Reserve Bank of Chicago publishes a US composite index based on a similar method but where only one factor of monthly data is chosen.[7] In our application, the number of factors was set equal to one. We also experimented with a larger numbers of factors, but the forecast performance was generally worse than in the one-factor model, see Appendix A.2 for details. The lag orders of the factor forecasting models in Eqs. (10), (11), and (13) were determined using the Bayesian Information Criterion (BIC).

3.3. Forecasting results

In Table 1, the out-of-sample forecasting results for German GDP are shown in terms of relative root mean squared forecast errors (RMSE), relative to the GDP standard deviation in the evaluation sample. Relative RMSEs smaller than one indicate informative forecasts; see Mittnik and Zadrozny (2005), for example. The forecasting results can be summarized as follows.

Comparison of IMS, DMS, and U forecasting methods. In part A of the table, we compare DMS with IMS and U forecasting, where all projections are based on the same estimated factors. The factors are estimated using real-time data, and the forecasts are compared with the final vintage GDP. For horizons up to 6 months, all three methods yield relative RMSEs smaller than one, indicating informative forecasts. For horizons four to six, the relative RMSEs are close to one, and thus have only a limited information content. Method U performs slightly better than the other methods at all horizons, whereas there are only very minor differences between the IMS and DMS forecasts.[8] Note that this result is in line with Boivin and Ng (2005, p. 148), where U forecasts performed best in a single-frequency framework. According to these findings, we only report U forecasts below.

Comparison with benchmark models. The forecast performance of the factor models is compared to that of simple benchmark models: a quarterly autoregressive (AR) model, which is estimated and forecast in real-time using IMS forecasting, and an AR model using DMS forecasting, see Marcellino et al. (2006) for details. Furthermore, we consider a naive forecast which is equal to the mean of the in-sample observations of GDP. The results can be found in part B of Table 1. Among the benchmarks, the in-sample mean performs no worse than the AR models. Both the AR model forecasts and the naive forecast are clearly outperformed by the factor models from part A of the table.

Role of the chosen vintage in the comparison. In part C of the table, we investigate the role of the reference vintage of the forecasts. In addition to the final vintage of GDP used above, the forecasts are evaluated by comparing them with the GDP, as obtained 3 and 12 months after the initial release of the GDP. It may happen that the final GDP is more

[6] Note that due to the publication lags of GDP, however, the effective forecast horizon needed for computing the forecasts has to be longer. For example, the data of vintage October 2004 (2004M10) contains GDP data up to 2004Q2 and monthly information up to 2004M9. For a forecast of the value in 2005Q1, we effectively need a three-quarter-ahead forecast from the end of the GDP sample.

[7] See Evans, Liu, and Pham-Kanter (2002) and Federal Reserve Bank of Chicago (2001).

[8] Note that the forecasts are not equal, but the differences are invisible for horizons larger than one, as the RMSEs were rounded to 2 digits after the decimal point.

Table 1
Out-of-sample forecast results, relative RMSE

Model	Dataset for estimation	GDP vintage for comparison	Monthly horizon					
			1	2	3	4	5	6
A. Factor forecasting with different forecast equations								
EM–IMS	Real-time	Final	0.85	0.84	0.89	0.93	0.94	0.96
EM–DMS	Real-time	Final	0.85	0.84	0.89	0.93	0.94	0.96
EM–U	Real-time	Final	0.85	0.84	0.86	0.89	0.91	0.90
B. Benchmark forecasts								
AR–IMS	Real-time	Final	1.01	1.03	1.02	1.03	1.03	1.03
AR–DMS	Real-time	Final	1.01	1.03	1.01	1.01	1.02	1.02
Mean	Real-time	Final	1.02	1.02	1.03	1.03	1.03	1.03
C. Alternative vintages for comparison								
EM–U	Real-time	3rd month	0.77	0.78	0.80	0.90	0.98	0.93
EM–U	Real-time	12th month	0.80	0.80	0.82	0.92	0.95	0.94
D. Forecast models estimated with data from alternative vintages								
EM–U	3rd month	3rd month	0.75	0.78	0.82	0.93	0.99	0.94
EM–U	12th month	12th month	0.78	0.82	0.86	0.95	1.00	0.99
EM–U	Final	Final	0.85	0.85	0.89	0.95	0.97	0.97
E. Forecasts without timely observations, forecasts with time-aggregated quarterly data								
EM–U	Real-time, no timely obs	Final	0.88	0.91	0.91	0.91	1.01	1.02
PCA–U	Real-time, quarterly	Final	0.95	0.99	1.00	1.00	1.01	1.02

Note: The entries represent root mean squared errors (RMSE) of the forecasts relative to the standard deviation of GDP growth in the evaluation sample. The model abbreviations are: EM is the EM algorithm together with PCA as in the appendix of Stock and Watson (2002a); IMS denotes iterative multi-step forecasting; DMS direct multi-step forecasting; and U unrestricted forecasting. PCA-U is unrestricted forecasting with factors estimated from time-aggregated quarterly data; and AR is an quarterly autoregressive model. The mean forecast is equal to the average of GDP growth computed over the subsample. The dataset for estimation is true real-time in line with the available data. Final data is from the latest vintage in the dataset available. 3rd and 12th month vintages refer to vintages available 3 or 12 months after the initial release of GDP, respectively. In panel E of the table, no timely data means cutting monthly observations until the last period in which GDP is available. Thus, there is no ragged-edge of the data, but the methods still have to tackle the mixed-frequency nature of the data. Quarterly data is obtained by time-aggregation of the respective vintage of the data.

affected by major revisions, perhaps benchmark revisions, which might make it difficult to obtain a good forecast performance. The results indicate that forecasts compared with GDP vintages 3 and 12 months after the initial release do indeed have smaller RMSEs. However, the decreases are rather small, in particular for horizons four to six.

Role of revisions. To discuss the importance of the revisions for forecasting, the forecast simulations were repeated using final vintage datasets for estimating the factors. Note that, in our final dataset, the timing and release conventions are consistent with those of the real-time database, as in Bernanke and Boivin (2003, p. 530), and delays in publication are incorporated as in the real-time dataset. Therefore, the two datasets differ only with respect to revisions in the data. For comparative purposes, we also consider different vintages, as in the previous subsection, to take into account the potential role of benchmarks revisions in the final vintage of data. Hence, for estimating the factor model we employ datasets where vintages released 3 and 12 months after the initial GDP release were used. The resulting forecasts are compared with the corresponding vintage of GDP. Thus, the forecast design eliminates the role of revisions 3 and 12 months after the initial release of GDP.

Part D of Table 1 contains the results from models estimated with the revised data, and can be compared with respective results in parts C and A, where unrevised data were used for estimation. We find that data revisions have no clear impact on the forecasting accuracy, as the use of final data leads to a performance similar to that with real-time data (comparing the third lines in parts D and A, respectively).[9] The results do not change when vintages 3 and 12 months after the initial GDP release are chosen as the data available for the estimation of the models (comparing the first and second lines in parts D and C, respectively). Overall, these results are in line with the findings of Bernanke and Boivin (2003), who conclude that revisions have only a small impact on forecasting US macroeconomic time series. Thus, data revisions cannot be blamed for the only moderate forecast performance of the factor models.

Data without timely monthly observations and aggregated quarterly data. In order to investigate the importance of the monthly observations that are available in a more timely fashion relative to GDP, another dataset is constructed, where observations of monthly data are available only up to the last month of the quarter for which GDP figures are available.[10] Hence, compared with the real-time dataset, timely observations from the monthly data are missing. We investigate whether it pays to exploit this information for factor forecasting. We also consider a quarterly factor model for forecasting and compare it with the results of the mixed-frequency factor model; see Mittnik and Zadrozny (2005) for a similar comparison in a different model framework. The dataset for the quarterly model can be obtained by time aggregation of the monthly time series for each vintage. Thus, together with the quarterly time series, the dataset is balanced, and there is no need to employ the EM algorithm. Note that time aggregation not only neglects the mixed-frequency nature of the data, but also implies a loss of information at the beginning or end of the sample. To forecast GDP, we employ unrestricted (U) forecasting similar to that in Eq. (13) above. Here, quarterly GDP is directly regressed on quarterly factors and their lags, with lag orders selected by the BIC.

Part E of the table provides the forecast results. When timely monthly observations are removed from the real-time data, the forecast performance worsens. The use of time-aggregated data for purely quarterly factor forecasting also leads to inferior results. It seems to be beneficial therefore to use the most timely information for forecasting GDP, and exploiting mixed-frequency data in contrast to time-aggregated data also seems to work slightly better in the present context. Note, however, that the advantages in relative RMSE of the EM factor models are small.

Relative forecast performance over time. Typically, a presentation of RMSEs alone gives no information about the models' forecast performance over time, but decision-makers might also require models with a relatively stable forecast accuracy period by period. We therefore investigate the relative performance of the models for every observation in the evaluation period. For this purpose, we present the forecasts, GDP growth, and squared forecast errors in Fig. 1.

A first impression of the forecast comparison is that the factor model forecasts are quite similar for most of the observations. Hence, in line with the average RMSE results from Table 1, there are no substantial differences between factor forecasts using real-time or final data, or by removing the most timely monthly observations or relying on a quarterly factor model only. The squared forecast errors shown in the bottom part of the figure decrease slightly from the beginning to the end of the evaluation sample. In the middle of the sample, the simple AR benchmark model is more clearly outperformed by the factor models than at the beginning or end. Hence, although the RMSE of the factor models is substantially smaller than the naive benchmarks according to Table 1, there seems to be no insurance against temporary relative forecast failure of the factor models in some periods of time.

4. Conclusions

In this paper, we employ factor models to forecast German GDP using mixed-frequency real-time data, where the time series are subject to different statistical publication lags. To cope with missing observations, we apply the EM algorithm together with principal components decomposition, following Stock and Watson (2002a). The EM algorithm provides us with monthly estimates of the factors, and we propose different ways of deriving iterative and direct forecasts for a low-frequency variable like GDP based on high-frequency factors.

To evaluate the short-term forecasting performance of the factor model, we carry out a recursive forecast comparison. The empirical application employs a medium-sized real-time dataset for post-unification Germany, which contains about fifty monthly and

[9] Note that in some cases the RMSE even increases when final data is used for factor estimation. The same result was found by Bernanke and Boivin (2003, Table 1, p. 532), who found that forecasting using final data often does not lead to a reduction in MSE compared with forecasting using real-time data.

[10] For a similar construction of a dataset, see Baffigi et al. (2004).

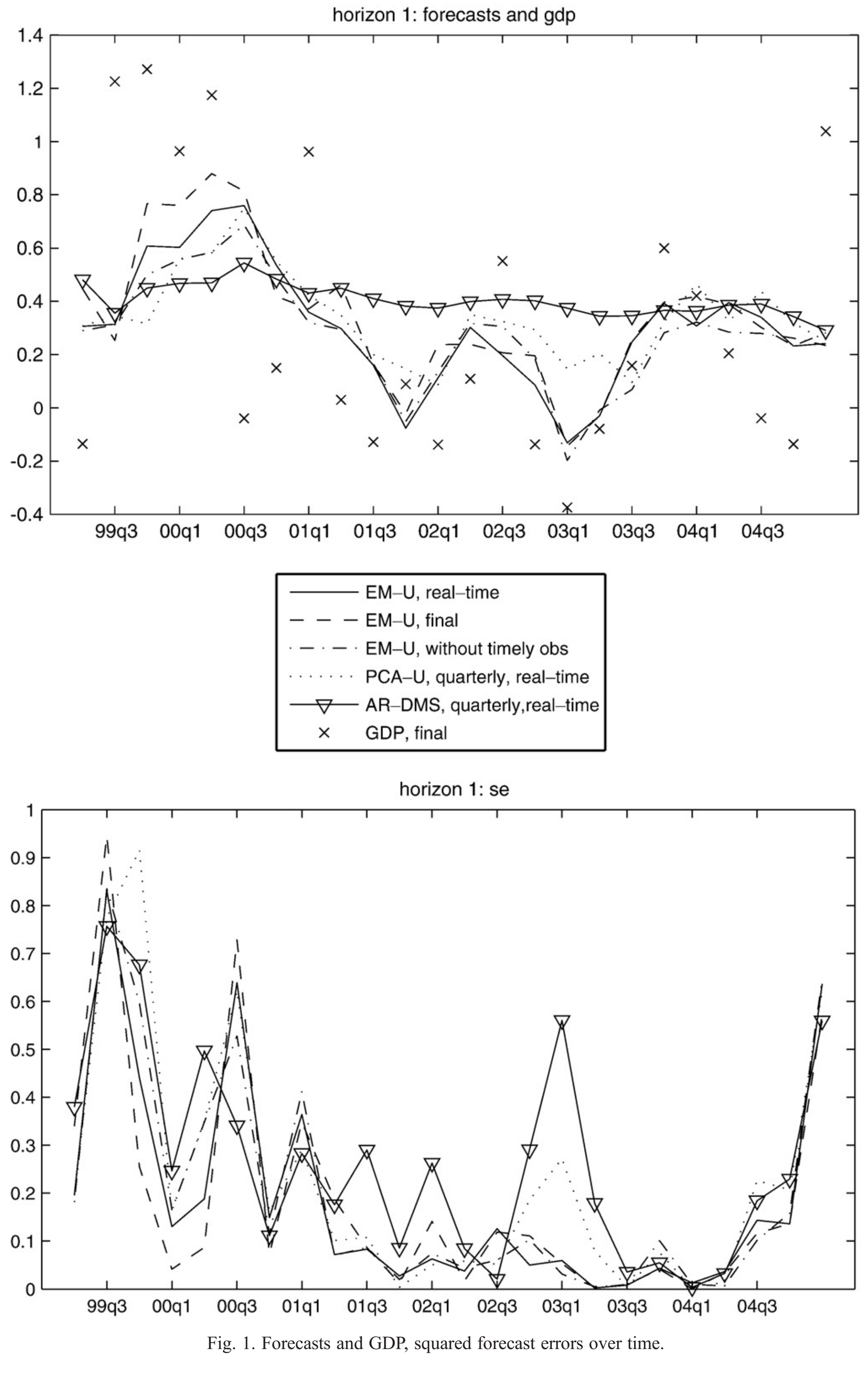

Fig. 1. Forecasts and GDP, squared forecast errors over time.

quarterly time series. Our empirical results indicate that data revisions have only a minor impact on the forecasting performance. Compared to factor models based on balanced data, the mixed-frequency factor model performs slightly better. More substantial differences arise when the real-time factor forecasts are compared with simple benchmark models. However, a comparison of forecast errors over time reveals that the differences between the factor model forecasts are relatively small.

Overall, the moderate forecast performance of the factor models indicates that forecasting German GDP in real-time is a difficult task. However, this is a finding in line with other work on the recent decline of forecast accuracy. For US macroeconomic time series in the 1990s, D'Agostino, Giannone, and Surico (2006), for example, find that simple benchmarks can hardly be outperformed by more sophisticated models. Similar results hold for the US Survey of Professional Forecasters, see Campbell (2007). Our results point in the same direction as these general results, and leave room for future research and further improvements of the models used here.

Since this is one of the first applications to this real-time dataset, a possible direction for future work could address a comparison with other methods for nowcasting and interpolating GDP at monthly intervals. Such an evaluation could provide an intuition for the relative performance of the method chosen here. Moreover, some sort of forecast combination among competing short-term monthly GDP forecasting models may be attractive due to a higher degree of robustness against outliers.[11]

Acknowledgements

This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank. We are grateful to Kai Carstensen, Jörg Döpke, Domenico Giannone, Hermann-Josef Hansen, Roman Liesenfeld, Ard den Reijer, and seminar participants at the Kiel Institute of World Economics, the DNB, the International Symposium on Forecasting ISF 2006 in Santander, the Econometric Society European Meeting ESEM 2006 in Vienna, the EUI Workshop on Factor Models in Theory and Practice 2006 in Florence, and the Bundesbank for helpful comments. The codes for this paper were written in Matlab. A number of functions for the forecast equations were taken from the Econometrics Toolbox written by James P. LeSage from www.spatial-econometrics.com.

Appendix A.1. Monthly and quarterly real-time dataset

The composite real-time dataset in this paper is an extension of the real-time dataset constructed by Gerberding, Worms, and Seitz (2005) and Gerberding, Kaatz, Worms, and Seitz (2005), which provides real-time vintages of German GDP and CPI inflation. The additional real-time data was taken from the Bundesbank Monthly Bulletin Supplement, Seasonal Adjusted Data. The release dates of the dataset are the same as the releases of the Bulletin Supplement, typically mid-month. The data releases available for the empirical application cover the period 1998M10 to 2005M6. Seasonal fluctuations in the data were eliminated by Bundesbank staff in real-time using Census-X12. In addition to the Bundesbank Statistical Bulletin data, we also consider monthly financial data and survey data. Overall, the dataset consists of 52 time series, 39 monthly series and 13 quarterly series that are described below in more detail.

Quarterly data. All quarterly time series are potentially subject to revisions, and thus vintages might differ between two releases, not only for the final observation.

1. GDP and components: gross domestic product; private consumption expenditure; government consumption expenditure; gross fixed capital formation: machinery and equipment; gross fixed capital formation: construction; exports; imports (7 series).
2. gross value added by sector: production sector excluding construction; hotels, restaurants and transport; financial, real estate renting and business services; public and private services (4 series).
3. income: gross wages and salaries; entrepreneurial and property income (2 series).

Monthly data. Of the 39 monthly time series, 13 are subject to revisions: industry statistics and the consumer price index. The remaining 26 monthly time series were assumed to be known immediately, and thus, in a particular month, we have observations of

[11] As an example of a forecast combination, see Kapetanios, Labhard, and Price (2008).

Table 2
Out-of-sample forecast results, relative RMSE

Model	Number of factors	Monthly horizon					
		1	2	3	4	5	6
EM–U	$r=1$	0.85	0.84	0.86	0.89	0.91	0.90
EM–U	$r=2$	0.97	1.03	0.94	0.97	0.97	0.96
EM–U	$r=3$	1.74	2.65	1.31	0.94	0.96	0.95

Note: The dataset for estimation and the GDP vintage for comparison are the same for all models. The dataset for estimation is real-time, whereas the GDP vintage is the final one. For details on the models and abbreviations, see Table 1 in the main text.

these time series for the previous month. These variables are assumed not to be subject to revisions.

1. industrial production: total manufacturing; total excluding construction; intermediate goods; capital goods; energy; durable consumer goods; non-durable consumer goods; construction (8 series, subject to revisions).
2. new orders received from the domestic economy: total manufacturing, capital goods; new orders received from abroad: total manufacturing, capital goods (4 series, subject to revisions).
3. consumer price index (subject to revisions).
4. interest rates: day-to-day money market rate; 1 month money market rate; 3 month money market rate; government bond yield 1 to 2 years maturity; government bond yield 9 to 10 years maturity; yield spread: government bonds 1 to 2 years maturity minus 3 months rate; yield spread: government bonds 9 to 10 years maturity minus 3 months rate (7 series).
5. share price indices: CDAX; DAX; REX German bond index (3 series).
6. exchange rates: US dollar/Deutsche Mark exchange rate; indicator of the German economy's price competitiveness against 19 industrial countries based on consumer prices (2 series).
7. surveys from IFO institute: business situation of producers of capital goods; producers of durable consumer goods; producers of non-durable consumer goods; retail trade; wholesale trade (5 series).
8. surveys from IFO institute: business expectations for the next 6 months of producers of capital goods; producers of durable consumer goods; producers of non-durable consumer goods; retail trade; wholesale trade (5 series).
9. surveys from IFO institute: stocks of finished goods of producers of capital goods; producers of durable consumer goods; producers of non-durable consumer goods (3 series).
10. consumer confidence index (survey GfK).

A.2. Forecasting results for different numbers of factors

As discussed in the main text, the forecasts obtained using the factor model are based on one factor. Below, we present forecasting results using up to three factors, using real-time data for estimating the factors and the final GDP for comparison. We report results for unrestricted (U) forecasting with real-time data only, as the results are very similar with other methods. In Table 2, the forecasting results are shown in terms of relative root mean squared forecast errors (RMSE).

As can be seen from the table, the forecast performance declines, as the number of factors is increased from one to three. If the number of factors is increased from two to three, the RMSE increases sharply. In more detail, the RMSE particularly increases for the nowcast (forecast horizons one to three), but not for longer horizons. This indicates that an over-large number of factors worsens the in-sample estimates in particular, for which timely monthly observations are used to extrapolate the monthly GDP. Furthermore, these results indicate that the EM algorithm fails to provide sensible estimates of monthly GDP if the number of factors is too large. This vulnerability is different to the single-frequency case, where the forecasting accuracy is often affected to a lesser extent by choosing a large number of factors; see, for example, the results from using a fixed number of factors in Stock and Watson (2002a), Tables 1–4. Following these results, we discuss the results in the main text for one factor.

References

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., & Veronese, G. (2006). New Eurocoin: Tracking economic growth in real time. *CEPR Working Paper* No. 5633.

Angelini, E., Henry, J., & Marcellino, M. (2006). Interpolation and backdating with a large information set. *Journal of Economic Dynamics & Control*, *30*, 2693−2724.

Baffigi, A., Golinelli, R., & Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, *20*, 447−460.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*, 191−221.

Bernanke, B., & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, *50*, 525−546.

Biernacki, C., Celeux, G., & Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, *41*, 561−575.

Boivin, J., & Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, *1*, 117−151.

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*, 169−194.

Campbell, S. (2007). Macroeconomic volatility, predictability, and uncertainty in the great moderation: Evidence from the survey of professional forecasters. *Journal of Business & Economic Statistics*, *25*, 191−200.

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure and mean–variance analysis in large asset markets. *Econometrica*, *51*, 1305−1324.

Chevillon, G., & Hendry, D. (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting*, *21*, 201−218.

Chow, G., & Lin, A. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *Review of Economics and Statistics*, *53*, 372−375.

Clements, M., & Galvão, A. (2007). *Macroeconomic forecasting with mixed frequency data: Forecasting US output growth.* Queen Mary University Working Paper No. 616.

D'Agostino, A., Giannone, D., & Surico, P. (2006). *(Un)Predictability and macroeconomic stability.* European Central Bank Working Paper No. 605.

Evans, M. (2005). Where are we now? Real-time estimates of the macroeconomy. *International Journal of Central Banking*, *1*, 127−175.

Evans, C., Liu, C. T., & Pham-Kanter, G. (2002). The 2001 recession and the Chicago Fed National Activity Index: Identifying business cycle turning points. *Federal Reserve Bank of Chicago Economic Perspectives*, *3Q/2002*, 26−43.

Federal Reserve Bank of Chicago (2001). *CFNAI background release.* www.chicagofed.org/economic research and data/files/cfnai background.pdf

Friedman, M. (1962). The interpolation of time series by related series. *Journal of the American Statistical Association*, *57*, 729−757.

Gerberding, C., Kaatz, M., Worms, A., & Seitz, F. (2005). A real-time data set for German key macroeconomic variables. *Schmollers Jahrbuch*, *125*, 337−346.

Gerberding, C., Worms, A., & Seitz, F. (2005). How the Bundesbank really conducted monetary policy: An analysis based on real-time data. *The North American Journal of Economics and Finance*, *16*, 277−292.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, *26*, 53−90.

Giannone, D., Reichlin, L., & Small, D. (2005). Nowcasting GDP and inflation: The real-time informational content of macroeconomic data releases. *Federal Reserve Board Finance and Economics Discussion Series 2005-42*.

Heaton, C., & Solo, V. (2006). Estimation of approximate factor models: Is it important to have a large number of variables? *Paper presented at the North American Summer Meetings of the Econometric Society 2006*.

Kapetanios, G., Labhard, V., & Price, S. (2008). Forecasting using Bayesian and information theoretic model averaging: An application to UK inflation. *Journal of Business & Economic Statistics*, *26*, 33−41.

Marcellino, M. (2007). Pooling-based data interpolation and backdating. *Journal of Time Series Analysis*, *28*, 53−71.

Marcellino, M., Stock, J., & Watson, M. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*, 499−526.

Mariano, R., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*, 427−443.

Mitchell, J., Smith, R., Weale, M., Wright, S., & Salazar, E. (2005). An indicator of monthly GDP and early estimate of quarterly GDP growth. *Economic Journal*, *115*, 108−129.

Mittnik, S., & Zadrozny, P. (2005). Forecasting quarterly German GDP at monthly intervals using monthly IFO business conditions data. In J. -E. Sturm & T. Wollmershaüser (Eds.), *IFO survey data in business cycle and monetary policy analysis* (pp. 19−48). Springer-Verlag.

Nunes, L. (2005). Nowcasting quarterly GDP growth in a monthly coincident indicator model. *Journal of Forecasting*, *24*, 575−592.

Sargent, T. (1987). *Dynamic macroeconomic theory.* Harvard University Press.

Stock, J., & Watson, M. (2002a). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*, 147−162.

Stock, J., & Watson, M. (2002b). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*, 1167−1179.