Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Factor-MIDAS for Now- and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP1

Massimiliano Marcellino and Christian Schumacher

Working Paper n. 333

20 November 2007

IGIER – Università Bocconi, Via Salasco 5, 20136 Milano –Italy
http://www.igier.uni-bocconi.it

The opinions expressed in the working papers are those of the authors alone, and not those of the Institute, which takes non institutional policy position, nor those of CEPR, NBER or Università Bocconi.

Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP[1]

Massimiliano Marcellino
Università Bocconi, IGIER and CEPR
massimiliano.marcellino@uni-bocconi.it

Christian Schumacher
Deutsche Bundesbank
christian.schumacher@bundesbank.de

First version: 3 October 2007, This version: 20 November 2007

[1]This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank. The paper has been presented at the Bank of England CCBS research forum 'New developments in dynamic factor modelling', the Bank of Canada workshop 'Forecasting Short-Term Developments and the Role of Econometric Models', and a seminar at the Bundesbank. We are grateful to the participants of these events as well as Sandra Eickmeier, Malte Knüppel, Gerhard Rünstler, Karsten Ruth, and Klaus Wohlrabe for helpful comments and discussions. The codes for this paper were written in Matlab. Some functions were taken from the Econometrics Toolbox written by James P. LeSage from `www.spatial-econometrics.com`. Other codes were kindly provided by Mario Forni from `www.economia.unimore.it/forni_mario/matlab.htm`, Arthur Sinko from `www.unc.edu/~sinko/midas.zip` and Gerhard Rünstler.

Abstract

This paper compares different ways to estimate the current state of the economy using factor models that can handle unbalanced datasets. Due to the different release lags of business cycle indicators, data unbalancedness often emerges at the end of multivariate samples, which is sometimes referred to as the 'ragged edge' of the data. Using a large monthly dataset of the German economy, we compare the performance of different factor models in the presence of the ragged edge: static and dynamic principal components based on realigned data, the Expectation-Maximisation (EM) algorithm and the Kalman smoother in a state-space model context. The monthly factors are used to estimate current quarter GDP, called the 'nowcast', using different versions of what we call factor-based mixed-data sampling (Factor-MIDAS) approaches. We compare all possible combinations of factor estimation methods and Factor-MIDAS projections with respect to nowcast performance. Additionally, we compare the performance of the nowcast factor models with the performance of quarterly factor models based on time-aggregated and thus balanced data, which neglect the most timely observations of business cycle indicators at the end of the sample. Our empirical findings show that the factor estimation methods don't differ much with respect to nowcasting accuracy. Concerning the projections, the most parsimonious MIDAS projection performs best overall. Finally, quarterly models are in general outperformed by the nowcast factor models that can exploit ragged-edge data.

JEL Classification Codes: E37, C53

Keywords: nowcasting, business cycle, large factor models, mixed-frequency data, missing values, MIDAS

1 Introduction

Many key indicators of macroeconomic activity are published by the statistical offices with a considerable time delay and at low frequencies. In particular, Gross Domestic Product (GDP) is typically published at quarterly frequency and has a considerable publication lag. In Germany, for example, GDP is released about five to six weeks after the end of the reference quarter. As policy makers regularly request information on the current state of the economy in terms of GDP, there is a need to provide estimates of current GDP in order to support policy decisions. For example, in April, German GDP is available only for the fourth quarter of the previous year. To obtain the current, second quarter GDP, we have to make a projection with forecast horizon of two quarters from the end of the GDP sample, using all currently available information in an efficient way. This projection is what we call the 'nowcast' in this paper, following, e.g., Giannone et al. (2005).

In general, it is difficult to exploit all information available for nowcasting, as business cycle indicators are released in an asynchronous way. Due to these different publication lags, multivariate datasets typically exhibit complicated patterns of missing values at the end of the sample and imply unbalanced samples for estimation. This leads to the so-called 'ragged-edge' data problem in econometrics, see Wallis (1986), and nowcast methods are necessary that can tackle this issue. Another difficulty arises, because GDP is quarterly data, whereas many important indicators are sampled at monthly or higher frequencies. Therefore, also a mixed-frequency problem has to be resolved for nowcasting.

In this paper, we discuss different ways to estimate factors from large high-frequency datasets subject to the ragged-edge problem, and how these factors can be used for nowcasting a low-frequency variable like GDP. In our description of the methods and the application below, factor nowcasting is essentially a two-step procedure, where factors are estimated in a first step, and the estimated factors enter specific projection models in a second step. Thus, according to the surveys in Boivin and Ng (2005), Eickmeier and Ziegler (2007), we follow the widely used two-step technique of factor forecasting, which is standard in case both the factors and the variable to be predicted are sampled at the same frequency.

For estimating the factors, we distinguish three main methods, which are all derived within a large scale dynamic factor model framework. First, we discuss the estimator by Altissimo et al. (2006), which builds upon the one-sided non-parametric dynamic principal component analysis (DPCA) factor estimator of Forni et al. (2005). To take into account the ragged-edge of the data, Altissimo et al. (2006) simply apply a realignment of each time series to obtain a balanced dataset. Second, we consider the Expectation-Maximisation (EM) algorithm combined with the factor estimator based static principal component analysis (PCA) as introduced by Stock and Watson (2002) and applied for forecasting and interpolation by Bernanke and Boivin (2003), Angelini, Henry and Marcellino (2006), and Schumacher and Breitung (2006). Third, we discuss the parametric state-space factor estimator of Doz, Giannone and Reichlin (2006), as applied in Giannone et al. (2005) and Banbura and Rünstler (2007).

Concerning the projection methods, we introduce the Factor-MIDAS approach. The starting point is the mixed-data sampling (MIDAS) framework proposed by Ghysels et al. (2004), and applied to macroeconomic variables in Clements and Galvão (2007). The basic MIDAS framework consists of a regression of a low frequency variable on a set of higher frequency indicators, where distributed lag functions are employed to specify the dynamic relationship. The Factor-MIDAS approach exploits estimated factors rather than single or small groups of economic indicators as regressors. Therefore, it directly translates the factor forecasting two-step approach as discussed in Boivin and Ng (2006) for the single-frequency case to the mixed-frequency case where factors are sampled at higher frequencies than the variable to be predicted. As in the standard MIDAS case, see Clements and Galvão (2007), direct multistep Factor-MIDAS forecasts are easily computed, which is convenient in our context.

We also evaluate a more general regression approach, where the dynamic relationship between the low frequency variable (GDP in our case) and the high frequency indicators (factors in our case) is unrestricted. This approach is based on the theoretical analysis in Marcellino and Schumacher (2007) and is labeled Factor-MIDAS-U, where U stands for unrestricted. As a third alternative, we consider a special regression scheme proposed by Altissimo et al. (2006), discuss how it can be used for nowcasting, and show its close relationship to the MIDAS method.

The main purpose of the paper is to compare empirically the different approaches of factor estimation in the presence of unbalanced data, combined with the alternative MIDAS projections. In particular, we apply the different methods to a large German dataset of about one hundred monthly indicators for nowcasting and short term forecasting German GDP growth. Germany is the largest country within the euro area, and this matters both from an economic and from a statistical point of view. For example, institutional forecasts for euro area macroeconomic variables by the Eurosystem and by the European Commission are often based on aggregation of the national forecasts. In these frameworks, Germany has a large weight, but existing nowcasts and forecasts for German GDP growth are not fully satisfying, see e.g. Schumacher and Breitung (2006). Furthermore, the quality of euro area data prior to 1999 is questionable, to the point that using German data prior to 1999 and euro area data afterwards can be preferable, see e.g. Lütkepohl and Brüggemann (2006).

In the empirical application, we evaluate the information content of nowcasts computed in each month of a given quarter, based on increasing information from the indicators. In addition, we investigate longer forecast horizons, up to two quarters ahead. In our recursive nowcast experiment, we consider the ragged-edge of the monthly data and the publication delay of GDP.

Furthermore, we discuss how the ragged-edge factor models perform compared with single-frequency factor models based on quarterly time-aggregated data. Quarterly data has been often used to forecast German GDP, see e.g. Schumacher (2007), and for other countries and datasets, for example by Marcellino et al. (2005) for Euro area countries' GDP using disaggregated and aggregated data, Banerjee et al. (2005) for euro area GDP, Banerjee and Marcellino (2006) for US GDP, Kapetanios et al. (2007) for UK GDP, and many others. Due to the widespread use of quarterly and partly time-aggregated data in the empirical literature, we will employ also factor forecasting based on time-aggregated quarterly data as a key benchmark for the nowcast factor models that can tackle ragged-edge data.

Additionally, we discuss the relative importance of static versus dynamic factors for nowcasting in our context. As there is some disagreement in the literature as to the appropriate estimation method of the factors, see Boivin and Ng (2005) and D'Agostino and Giannone (2006), a nowcast comparison should also address this issue.

Finally, since some of the factor estimation methods discussed above allow for an integrated approach of estimating the factors and nowcasting in one single step, in particular the state-space approach by Giannone et al. (2005) and Banbura and Rünstler (2007), we compare our two-step Factor-MIDAS procedure with the integrated approach.

The paper is structured as follows. Section 2 reviews the competing approaches to factor nowcasting under analysis, and the different MIDAS projection methods. Section 3 presents the empirical nowcast exercise, and compares and discusses the results. Section 4 summarises and concludes.

2 Factor nowcasting with ragged-edge data

In this paper we focus on quarterly GDP growth, which is denoted as y_{t_q} where t_q is the quarterly time index $t_q = 1, 2, 3, \ldots, T_q$. GDP growth can also be expressed at the monthly frequency by setting $y_{t_m} = y_{t_q} \forall t_m = 3t_q$ with t_m as the monthly time index. Thus, GDP y_{t_m} is observed only at months $t_m = 3, 6, 9, \ldots, T_m$ with $T_m = 3T_q$. The aim is to nowcast or forecast GDP h_q quarters ahead, or $h_m = 3h_q$ months ahead, based on information in month T_m, denoted as $y_{T_m+h_m|T_m}$. For example, since GDP for the first quarter of a given year is released around mid-May, a nowcast can be produced in January, February, and March of the current year, while a forecast can be produced in any month of the previous year.

The information set includes a large set of stationary monthly indicators, collected in the N-dimensional vector $\mathbf{X}_{t_m}$. The time index t_m denotes monthly frequency and $\mathbf{X}_{t_m}$ is fully available for each month $t_m = 1, 2, 3, \ldots, T_m$. However, due to publication lags, some elements at the end of the sample can be missing, thus rendering an unbalanced sample of $\mathbf{X}_{t_m}$.

We want to model $\mathbf{X}_{t_m}$ using a dynamic factor specification, and use the estimated factors, which efficiently summarize the information in $\mathbf{X}_{t_m}$, to nowcast and forecast GDP growth, y_{T_q}. According to Boivin and Ng (2005), factor forecasting with large, single-frequency datasets is often carried out using a similar two-step procedure: Firstly, the factors are estimated, and secondly, a dynamic model for the variable to be predicted is augmented with the estimated factors, see Bai and Ng (2006) for technical details on the properties of the resulting forecasts. However, to take into account the specific nowcast framework, the following modifications are necessary:

1. The first step factor estimation methods have to be able to handle ragged-edge data, due to the missing values at the end of the sample in a real time context.

2. The second step regression methods have to be able to handle mixed frequency data, in particular a low-frequency target variable and higher-frequency factors.

We will firstly discuss the proper factor estimation methods in subsection 2.1, and then the factor based nowcast regression methods in subsection 2.2.[1]

2.1 Estimating the factors with ragged-edge data

We assume that the monthly observations have a factor structure according to

$$\mathbf{X}_{t_m} = \mathbf{\Lambda}\mathbf{F}_{t_m} + \boldsymbol{\xi}_{t_m}, \tag{1}$$

where the r-dimensional factor vector is denoted as $\mathbf{F}_{t_m} = (f'_{1,t_m}, \ldots, f'_{r,t_m})'$. The factors times the $(N \times r)$ loadings matrix $\mathbf{\Lambda}$ represent the common components of each variable. The idiosyncratic components $\boldsymbol{\xi}_{t_m}$ are that part of $\mathbf{X}_{t_m}$ not explained by the factors.

Under the assumption that the $(T_m \times N)$ data matrix $\mathbf{X}$ is balanced, various ways to estimate the factors have been provided in the literature. For example, two of the most widely used approaches are based on PCA as in Stock and Watson (2002) or dynamic PCA according to Forni et al. (2005). For overviews, see the surveys by Stock and Watson (2006), section 4, and Boivin and Ng (2005) and the comparisons by D'Agostino and Giannone (2006) and Schumacher (2007). Note that, according to (1), all the factor models to be discussed below will work at the higher monthly frequency, thus factor estimates are available for all monthly observations $t_m = 1, 2, 3, \ldots, T_m$.

Vertical realignment of data and dynamic principal components factors A very convenient way to solve the ragged-edge problem is provided by Altissimo et al. (2006) for estimating the New Eurocoin indicator. They propose to realign each time series in the sample in order to obtain a balanced dataset, see also Schneider and Spitzer (2004). Assume that variable i is released with k_i months of publication lag. Thus, given a dataset in period T_m, the final observation

[1] To focus on ragged-edge and mixed-frequency problems, we abstract from additional complications such as those resulting from seasonal adjustment and data revisions.

available of this time series is for period $T_m - k_i$. The realignment proposed by Altissimo et al. (2006) is then simply

$$\widetilde{x}_{i,T_m} = x_{i,T_m - k_i} \tag{2}$$

for $t_m = k_i + 1, \ldots, T_m$. Applying this procedure for each series, and harmonising at the beginning of the sample, yields a balanced data set $\widetilde{\mathbf{X}}_{t_m}$ for $t_m = \max(\{k_i\}_{i=1}^N) + 1, \ldots, T_m$.

Given this monthly data, Altissimo et al. (2006) propose to use dynamic PCA to estimate the factors. As the dataset is balanced, the two-step estimation techniques by Forni et al. (2005) directly apply. In our applications below, we will denote the combination of vertical realignment and dynamic principal components factors as 'VA-DPCA'.

The vertical realignment solution to the ragged-edge problem is easy to use. A disadvantage is that the availability of data determines dynamic cross-correlations between variables. Furthermore, statistical release dates for data are not the same over time, for example, due to major revisions. In this case, dynamic correlations within the data change and factors can change over time. The same holds if factors are reestimated at a higher frequency than the frequency of the factor model. This is a very common scenario, for example, if a monthly factor model is reestimated several times within a month when new monthly observations are released. If this the case, the realignment of the data changes the correlation structure all the time. On the other hand, dynamic PCA as in Forni et al. (2005) exploits the dynamic cross-correlations in the frequency domain and might be in principle able to account for these changes in realignments of the data.

Principal components factors and the EM algorithm To consider missing values in the data for estimating factors, Stock and Watson (2002) propose an EM algorithm together with the standard PCA. Consider a variable i from the dataset $\mathbf{X}_{t_m}$ as a full data column vector $\mathbf{X}_i = (x_{i,1}, \ldots, x_{i,T_m})'$. Assume that not all the observations are available due to the ragged-edge problem. The vector $\mathbf{X}_i^{\text{obs}}$ contains the observations available for variable i, which is only a subset of $\mathbf{X}_i$ due to missing values. We can formulate the relationship between observed and not fully observed data by

$$\mathbf{X}_i^{\text{obs}} = \mathbf{A}_i \mathbf{X}_i, \tag{3}$$

where $\mathbf{A}_i$ is a matrix that can tackle missing values or mixed frequencies. In case no observations are missing, $\mathbf{A}_i$ is the identity matrix. In case an observation is missing at the end of the sample, the corresponding final row of the identity matrix is removed to ensure (3). The EM algorithm proceeds as follows:

1. Provide an initial (naive) guess of observations $\widehat{\mathbf{X}}_i^{(0)} \forall i$. These guesses together with the fully observable monthly time series yields a balanced dataset $\widehat{\mathbf{X}}^{(0)}$. Standard PCA provides initial monthly factors $\widehat{\mathbf{F}}^{(0)}$ and loadings $\widehat{\mathbf{\Lambda}}^{(0)}$.

2. **E-step:** An update estimate of the missing observations for variable i is provided by the expectation of $\mathbf{X}_i$ conditional on observations $\mathbf{X}_i^{\text{obs}}$, factors $\widehat{\mathbf{F}}^{(j-1)}$ and loadings $\widehat{\mathbf{\Lambda}}_i^{(j-1)}$ from the previous iteration

$$\widehat{\mathbf{X}}_i^{(j)} = \widehat{\mathbf{F}}^{(j-1)}\widehat{\mathbf{\Lambda}}_i^{(j-1)} + \mathbf{A}_i'(\mathbf{A}_i'\mathbf{A}_i)^{-1}\left(\mathbf{X}_i^{\text{obs}} - \mathbf{A}_i\widehat{\mathbf{F}}^{(j-1)}\widehat{\mathbf{\Lambda}}_i^{(j-1)}\right). \tag{4}$$

   The update consists of two components: the common component from the previous iteration $\widehat{\mathbf{F}}^{(j-1)}\widehat{\mathbf{\Lambda}}_i^{(j-1)}$, plus the low-frequency idiosyncratic component $\mathbf{X}_i^{\text{obs}} - \mathbf{A}_i\widehat{\mathbf{F}}^{(j-1)}\widehat{\mathbf{\Lambda}}_i^{(j-1)}$, distributed by the projection coefficient $\mathbf{A}_i'(\mathbf{A}_i'\mathbf{A}_i)^{-1}$ on the high-frequency periods. For general issues see Stock and Watson (2002), and for a discussion of the properties in the ragged-edge case, see Schumacher and Breitung (2006).

3. **M-step:** Repeat the E-step for all i yielding again a balanced dataset. Reestimate the factors and loadings, $\widehat{\mathbf{F}}^{(j)}$ and $\widehat{\mathbf{\Lambda}}^{(j)}$ by PCA, and go to step 2 until convergence.

After convergence, the EM algorithm provides monthly factor estimates $\widehat{\mathbf{F}}_{t_m}$ as well as estimates of the missing values of the time series. Thus, interpolation of missing values as well as factor estimation is carried out consistently in the factor framework (1) with factors estimated by PCA. For a detailed discussion of the properties of the EM algorithm for interpolation and backcasting, see Angelini et al. (2006). In the applications below, we will denote the this factor estimator as 'EM-PCA'.

Estimation of a large parametric factor model in state-space form The approach followed by Doz et al. (2006) and Kapetanios and Marcellino (2006) casts the large factor model in state-space form. However, Kapetanios and Marcellino (2006) estimate the factors using subspace algorithms, while Doz et al. (2006) exploit the Kalman filter and smoother. Here, we follow the Doz et al. (2006) approach as it can be more directly applied to ragged-edge data, see Giannone et al. (2005).

To specify a complete model, an explicit dynamic VAR structure is assumed to hold for the factors. The full state-space model has the form

$$\mathbf{X}_{t_m} = \mathbf{\Lambda}\mathbf{F}_{t_m} + \boldsymbol{\xi}_{t_m}, \tag{5}$$

$$\Psi(L_m)\mathbf{F}_{t_m} = \mathbf{B}\boldsymbol{\eta}_{t_m}. \tag{6}$$

Equation (5) is the static factor representation of $\mathbf{X}_{t_m}$ as above in (1). Equation (6) specifies a VAR of the factors with lag polynomial $\Psi(L_m) = \sum_{i=1}^{p}\Psi_i L_m^i$ and L_m is the monthly lag operator with $L_m x_{t_m} = x_{t_m-1}$. The q-dimensional vector $\boldsymbol{\eta}_{t_m}$ contains the orthogonal dynamic shocks that drive the r factors, where the matrix $\mathbf{B}$ is $(r \times q)$-dimensional. The model is already in state space form, since the factors $\mathbf{F}_{t_m}$ are the states. If the dimension of $\mathbf{X}_{t_m}$ is small, the model can be estimated using ML. In order to account for large datasets, Doz et al. (2006) propose quasi-ML to estimate the factors, as iterative ML is infeasible in this framework. For a given number of factors r and dynamic shocks q, the estimation proceeds in the following steps:

1. Estimate $\widehat{\mathbf{F}}_{t_m}$ using PCA as an initial estimate.

2. Estimate $\widehat{\mathbf{\Lambda}}$ by regressing $\mathbf{X}_{t_m}$ on the estimated factors $\widehat{\mathbf{F}}_{t_m}$. The covariance of the idiosyncratic components $\widehat{\boldsymbol{\xi}}_{t_m} = \mathbf{X}_{t_m} - \widehat{\mathbf{\Lambda}}\widehat{\mathbf{F}}_{t_m}$, denoted as $\widehat{\mathbf{\Sigma}}_{\boldsymbol{\xi}}$, is also estimated.

3. Estimate factor VAR(p) on the factors $\widehat{\mathbf{F}}_{t_m}$ yielding $\widehat{\Psi}(L)$ and the residual covariance of $\widehat{\boldsymbol{\varsigma}}_{t_m} = \widehat{\Psi}(L_m)\widehat{\mathbf{F}}_{t_m}$, denoted as $\widehat{\mathbf{\Sigma}}_{\boldsymbol{\varsigma}}$.

4. To obtain an estimate for $\mathbf{B}$, given the number of dynamic shocks q, apply an eigenvalue decomposition of $\widehat{\mathbf{\Sigma}}_{\boldsymbol{\varsigma}}$. Let $\mathbf{M}$ be the $(r \times q)$-dimensional matrix of the eigenvectors corresponding to the q largest eigenvalues, and let the $(q \times q)$-dimensional matrix $\mathbf{P}$ contain the largest eigenvalues on the main diagonal and zero otherwise. Then, the estimate of $\mathbf{B}$ is $\widehat{\mathbf{B}} = \mathbf{M} \times \mathbf{P}^{-1/2}$.

5. The coefficients and auxiliary parameters of the system of equations (5) and (6) is fully specified numerically. The model is cast into state-space form. The Kalman filter or smoother then yield new estimates of the monthly factors.[2]

If missing values at the end of the sample are present, as in our setup, the Kalman filter also yields optimal estimates and forecasts conditional on the model structure and properties of the shocks. Thus, it is well suited to tackle ragged-edge problems as in the present context. Nonetheless, one has to keep in mind that in this case the coefficients in system matrices have to be estimated from a balanced sub-sample of data, as in step 1 a fully balanced dataset is needed for PCA initialisation. However, although the system matrices are estimated on balanced data in the first step, the factor estimation based on the Kalman filter applies to the unbalanced data and can tackle ragged-edge problems. The solution is to estimate coefficients outside the state-space model and avoid estimating a large number of coefficients by iterative ML.

In comparison with the EM algorithm discussed above, the state-space estimation also considers dynamics of the factors explicitly, whereas the static factor models doesn't. In the applications below, we will denote the state-space model Kalman filter estimator of the factors as 'KFS-PCA'.

[2] It is worth mentioning that when the model parameters are estimated using factors obtained by subspace algorithms, as in Kapetanios and Marcellino (2006), simulation experiments indicate that the Kalman filter based factors are very close to the original subspace factors.

2.2 Nowcasting and forecasting quarterly GDP with Factor-MIDAS

To forecast quarterly GDP using the estimated monthly factors, we rely on the mixed-data sampling (MIDAS) approach as proposed by Ghysels and Valkanov (2006), Ghysels et al. (2007), Clements and Galvão (2007), and Marcellino and Schumacher (2007). The MIDAS regression approach is a direct forecasting tool, as no dynamics on the factors nor joint dynamics for GDP and the factors are explicitly modelled. Rather, MIDAS forecasts directly relate future GDP to current and lagged indicators, thus yielding different forecast models for each forecast horizon, see Marcellino, Stock and Watson (2006) as well as Chevillon and Hendry (2005) for detailed discussions of this issue in the single-frequency case.

The basic Factor-MIDAS approach In the standard MIDAS approach economic variables at higher frequency are used as regressors, while in our Factor-MIDAS the explanatory variables are estimated factors. Let us assume for simplicity that we have only one factor $\widehat{f}_{t_m}$ for forecasting and $r = 1$. Hence, the forecast model for forecast horizon h_q quarters with $h_q = h_m/3$ is

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \beta_1 b(L_m, \boldsymbol{\theta}) \widehat{f}_{t_m}^{(3)} + \varepsilon_{t_m+h_m}, \tag{7}$$

where the polynomial $b(L_m, \boldsymbol{\theta})$ is the exponential Almon lag with

$$b(L_m, \theta) = \sum_{k=0}^{K} c(k, \boldsymbol{\theta}) L_m^k, \quad c(k, \boldsymbol{\theta}) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=0}^{K} \exp(\theta_1 k + \theta_2 k^2)}. \tag{8}$$

In the MIDAS approach, quarterly GDP $y_{t_q+h_q}$ is directly related to the factor $\widehat{f}_{t_m}^{(3)}$ and its lags, where $\widehat{f}_{t_m}^{(3)}$ is a skip-sampled version of the monthly factor $\widehat{f}_{t_m}$ as estimated in the sections above. The superscript three indicates that every third observation starting from the t_m-th one is included in the regressor $\widehat{f}_{t_m}^{(3)}$, thus $\widehat{f}_{t_m}^{(3)} = \widehat{f}_{t_m} \,\forall\, t_m = \ldots, T_m - 6, T_m - 3, T_m$. Lags of the monthly factors are treated accordingly, e.g. the k-th lag $\widehat{f}_{t_m-k}^{(3)} = \widehat{f}_{t_m-k} \,\forall\, t_m = \ldots, T_m - k - 6, T_m - k - 3, T_m - k$.

For given $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$, the exponential lag function $b(L_m, \boldsymbol{\theta})$ provides a parsimonious way to consider monthly lags of the factors as we can allow for large K to approximate the impulse response function of GDP from the factors. The longer the lead-lag relationship in the data is, the less MIDAS suffers from sampling uncertainty compared with the estimation of unrestricted lags, where the number of coefficients increases with the lag length.

The MIDAS model can be estimated using nonlinear least squares (NLS) in a regression of y_{t_m} onto $\widehat{f}_{t_m-k}^{(3)}$, yielding coefficients $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\widehat{\beta}_0$ and $\widehat{\beta}_1$. The forecast is given by

$$y_{T_m+h_m|T_m} = \widehat{\beta}_0 + \widehat{\beta}_1 b(L_m, \widehat{\boldsymbol{\theta}}) \widehat{f}_{T_m}. \tag{9}$$

For the case of $r > 1$ with $\mathbf{F}_{t_m} = (f'_{1,t_m}, \ldots, f'_{r,t_m})'$, the model generalises to

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \sum_{i=1}^{r} \beta_{1,i} b_i(L_m, \boldsymbol{\theta}_i) \widehat{f}^{(3)}_{i,t_m} + \varepsilon_{t_m+h_m}. \tag{10}$$

Here, the parameters $\boldsymbol{\theta}_i$, that determine the curvature of the impulse response function, can vary between the different factors. The estimation and forecast is otherwise the same.

Since all our applications are factor based, we drop the prefix 'Factor' and denote this approach as 'MIDAS-basic'.

Smoothed MIDAS Another way to formulate a mixed-frequency projection is employed in the New Eurocoin index, see Altissimo et al. (2006). New Eurocoin is a composite indicator of the Euro area economy and can be regarded as a projection of smoothed GDP on monthly factors, see Altissimo et al. (2006), section 4. Although the methods in that paper aim at deriving a composite coincident indicator, not explicitly now- or forecasts, one can directly generalise them for these purposes.

In particular, the projection can be written as

$$y_{T_m+h_m|T_m} = \widehat{\mu} + \mathbf{G}\widehat{\mathbf{F}}_{T_m}, \tag{11}$$

$$\mathbf{G} = \widetilde{\boldsymbol{\Sigma}}_{y\mathbf{F}}(h_m) \times \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{-1}, \tag{12}$$

where $\widehat{\mu}$ is the sample mean of GDP, assuming that the factors have mean zero, and $\mathbf{G}$ is a projection coefficient matrix. $\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}$ is the estimated sample covariance of the factors, and $\widetilde{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$ is a particular cross-covariance with k monthly lags between GDP and the factors. The tilde denotes that $\widetilde{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$ is not an estimate of the sample cross-covariance between factors and GDP, rather a cross-covariance between smoothed GDP and factors. The smoothing aspect is introduced into $\widetilde{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$ as follows: Assume that both the factors and GDP are demeaned. Then, let the covariance between $\widehat{\mathbf{F}}_{t_m-k}$ and y_{t_m} be estimated by

$$\widehat{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k) = \frac{1}{T^* - 1} \sum_{t_m=M+1}^{T_m} y_{t_m} \widehat{\mathbf{F}}^{(3)\prime}_{t_m-k}, \tag{13}$$

where $T^* = \text{floor}[(T_m - (M+1))/3]$ is the number of observations available to compute the cross-covariances for $k = -M, \ldots, M$ and $M \geq 3h_q = h_m$. Note that skip-sampled factors $\widehat{\mathbf{F}}^{(3)\prime}_{t_m-k}$ enter $\widehat{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$, as we have only quarterly observations of GDP. Given $\widehat{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$, we can estimate the cross-spectral matrix

$$\widehat{\mathbf{S}}_{y\mathbf{F}}(\omega_j) = \sum_{k=-M}^{M} \left(1 - \frac{|k|}{M+1}\right) \widehat{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k) e^{-i\omega_j k} \tag{14}$$

at frequencies $\omega_j = \frac{2\pi j}{2H}$ for $i = -H, \ldots, H$ using a Bartlett lead-lag window. The low-frequency relationship between $\widehat{\mathbf{F}}_{t_m-k}$ and y_{t_m} in New Eurocoin is obtained by filtering out cross fluctuations at frequencies larger than $\pi/6$, using the frequency-response function $\alpha(\omega_j)$, which is defined as $\alpha(\omega_j) = 1 \,\forall\, |\omega_j| < \pi/6$ and zero otherwise. By inverse Fourier transform we obtain the autocovariance matrix $\widetilde{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$ reflecting low-frequency comovements between $\widehat{\mathbf{F}}_{t_m-k}$ and y_{t_m}

$$\widetilde{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k) = \frac{1}{2H+1} \sum_{j=-H}^{H} \alpha(\omega_j) \widehat{\mathbf{S}}_{y\mathbf{F}}(\omega_j) e^{i\omega_j k}, \tag{15}$$

which is part of the projection coefficients (12) for $k = 1, 2, \ldots, h_m = 3h_q$ months. For given M and H, we can compute the projection (11). We will denote this MIDAS approach as 'MIDAS-smooth'.

The relationship between the basic MIDAS approach in (7) or (10) and MIDAS-smooth is immediately clear when we disregard the smoothing aspect for a moment, and consider $\widehat{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$ instead of $\widetilde{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$ in the projection coefficient $\widehat{\boldsymbol{\Sigma}}_{y\mathbf{F}}(h_m) \times \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{-1}$ in (12). First note that $\widehat{\boldsymbol{\Sigma}}_{y\mathbf{F}}(k)$ is a consistent estimator of the true cross-covariance, if the sample size is sufficiently large, despite the missing values. MIDAS-basic (7) and its multivariate extension (10) are based on the same finding as the smooth projection: one regresses low-frequency GDP on skip-sampled high-frequency factors, but with a different functional (exponential lag) form and allows for non-zero lag orders. Thus, in terms of lags considered, the New Eurocoin projection is a restricted form of MIDAS-basic, but with a different weighting.

The unrestricted MIDAS The MIDAS-basic and MIDAS-AR rely on the exponential lag function, whereas MIDAS-smooth considers only t_m-dated factors as regressors in a particular way. As an alternative to these approaches, we also consider an unrestricted lag order model

$$y_{t_m+h_m} = \beta_0 + \mathbf{D}(L_m)\widehat{\mathbf{F}}_{t_m}^{(3)} + \varepsilon_{t_m+h_m}, \tag{16}$$

where $\mathbf{D}(L_m) = \sum_{k=0}^{K} \mathbf{D}_k L_m^k$ is an unrestricted lag polynomial of order K. A theoretical justification for this specification is provided in Marcellino and Schumacher (2007), who show that it can be derived as an approximation to the model resulting from mixed sampling from a higher frequency ARMA model.

We estimate $\mathbf{D}(L_m)$ and β_0 by OLS. To specify the lag order in the empirical application, we consider a fixed scheme with $k = 0$ and an automatic lag length selection using the Bayesian information criterion (BIC). Note that for $k = 0$, we consider only t_m-dated factors for forecasting. Thus, with $k = 0$ the projection model is close to the MIDAS-smooth projection as employed in the New Eurocoin index, see the discussion above. The difference is of course that the smoothing aspect is neglected here.

The unrestricted MIDAS with $k = 0$ can be regarded as the most simple form of MIDAS, and can serve as a benchmark against the more distinguished alternatives above. We will denote the unrestricted MIDAS with $k = 0$ as 'MIDAS-U0', and with estimated lag order by BIC as 'MIDAS-U'.

3 Empirical nowcast and forecast comparison

The empirical application will be carried out in a recursive nowcast experiment. In subsection 3.1, we describe the design of this exercise, the data used and the specifications of the models. In the following subsections, the empirical results for German GDP nowcasting will be discussed. In particular, following the methodological discussion above, we present

- a comparison of factor estimation methods that can tackle ragged-edge data in section 3.2, and
- a comparison of MIDAS projections in section 3.3.

To relate our results to earlier empirical findings and conceptual discussions in the factor forecast literature, further results are provided:

- Section 3.4: A comparison of monthly nowcast models with quarterly factor models,
- Section 3.5: A discussion of static versus dynamic factor estimation, and
- Section 3.6: A comparison of the two-step nowcast approach chosen here with an integrated state-space model.

3.1 Design of the nowcast and forecast comparison exercise

Data and replication of the ragged edge The dataset contains German quarterly GDP from $1992Q1$ until $2006Q3$ and 111 monthly indicators from $1992M1$ until $2006M11$. The dataset is a final dataset. It is not a real-time dataset and does not contain vintages of data, as they are not available for Germany for such a broad coverage of time series. Furthermore, in Schumacher and Breitung (2006), a considerably smaller real-time dataset for Germany is used, but the results indicate that data revisions do not affect the forecast accuracy considerably. Similar results have been found by Boivin and Ng (2003) for the US in a similar context. More information about the data can be found in appendix A.

To consider the ragged-edge of the data at the end of the sample due to different publication lags, we follow Banbura and Rünstler (2007) and replicate the ragged-edge from the one final vintage of data that is available. When downloading the data - the download date for the data used here was 6th December 2006 -, we observe the ragged-edge pattern in terms of the missing values at the end of the data sample. For example, at the beginning of December 2006, we observe interest rates until November 2006, thus there is only one missing value at the end of the sample, whereas industrial production is available up to September 2006, implying three missing values. For each time series, we store the missing values at the end of the sample. Under the assumption that these patterns of data availability remain stable over time, we can impose the same missing values at each point in time of the recursive experiment. Thus, we shift the missing values back in time to mimic the availability of information as in real time.

Nowcast and forecast design To evaluate the performance of the models, we carry out recursive estimation and nowcasting, where the full sample is split into an evaluation sample and an estimation sample, which is recursively expanded over time. The evaluation sample is between $1998Q4$ and $2006Q3$. For each of these quarters, we want to compute nowcasts and forecasts depending on different monthly information sets. For example, for the initial evaluation quarter $1998Q4$, we want to compute a nowcast in December 1998, one in November, and October, whereas the forecasts are computed from September 1998 backwards in time accordingly. Thus, we have three nowcasts computed at the beginning of each of the intra-quarter months. Concerning the forecasts, we present results up to two quarters ahead. Thus, again for the initial evaluation quarter $1998Q4$, we have six forecasts computed based on information available in April 1998 up to information available in September 1998. Overall, we have nine projections for each GDP observation of the evaluation period, depending on the information available to make the projection.

The estimation sample depends on the information available at each period in time when computing the now- and forecasts. Assume again we want to nowcast GDP for $1998Q4$ in December 1998, then we have to identify the time series observations available at that period in time. For this purpose, we exploit the ragged-edge structure from the end of the full sample of data, as discussed in the previous subsection. For example, for the nowcast GDP for $1998Q4$ made in December 1998, we know from our full sample that at each period in time, we have one missing value for interest rates and three missing values of industrial production. These missing values are imposed also for the period December 1998, thus replicating the same ragged-edge pattern of data availability. We do this accordingly in every recursive subsample to determine the pseudo real-time final observation of each time series. The first observation for each time series is the same for all recursions, namely $1992M1$. This implies the recursive design with increasing information over time available for estimating the factor models. To replicate the publication lags of GDP, we exploit the fact that GDP of the previous quarter is available for now- and forecasting at the beginning of the third month of the next quarter. Note that we reestimate the factors and forecast equations every recursion when new information becomes available, so factor weights and forecast model coefficients are allowed to change over time.

For each evaluation period, we compute nine now- and forecasts depending on the available information. To compare the nowcasts with the realisations of GDP growth, we use the mean-squared error (MSE). As a measure of informativeness of the nowcasts, we relate the MSE to the variance of GDP, where the variance is computed over the evaluation period, see Forni et al. (2003). A relative MSE to GDP variance less than one indicates that the forecast of a model for the chosen now- and forecast horizon is to some extent informative for current and future GDP. Note that this relative statistic can also be interpreted as a measure to compare the MSE of the factor models with the corresponding MSE of the out-of-sample mean of GDP as a naive forecast.

Specification of factor models To specify the number of factors in the applications below, we follow two approaches: We determine the number of static and dynamic factors, r and q, respectively, using information criteria from Bai and Ng (2002) and Bai and Ng (2007). Additionally, we compute now- and forecasts for all possible combinations of r and q and evaluate them. In our application, we consider a maximum of $r = 6$ and all combinations of r and q with $q \leq r$. Details can be found in the appendix B. The key result from this exercise is that only for the case $r = 1$ and partly for $r = 2$, now- and forecasts have information content for current and future GDP. Apart from a few exceptions, all other combinations of numbers of factors - including those determined by information criteria - performed worse than the specifications we provide results for in the main text below. A plausible reason for this result is the combination of the rather short estimation sample and the substantial likelihood of parameter changes. In this case, Banerjee, Marcellino and Masten (2007) show that there is a substantial deterioration in the performance of forecasts based on many factors, and model specification by information criteria is not helpful. Also for the US, it was shown that only very few factors can obtain satisfactory forecast results, see Stock and Watson (2002). Due to these findings and to preserve space, we only present results for $r = 1$ below. We don't present results for $r = 2$, as the main results and conclusions are the same.

For estimating the state-space factor model, a lag order determination is required to specify the factor VAR(p). For this purpose, we apply the Bayesian information criterion (BIC) with a maximum lag order of $p = 6$ months. The chosen lag lengths are usually very small with only one or two lags in most of the cases. To specify the dynamic PCA estimator and MIDAS-smooth, we use the frequency-domain parameters $M = 24$ and $H = 60$ for estimating the spectral density.

The EM algorithm we implement for monthly factor estimation is slightly different from what described above. In particular, we do not update the factor weights during the iterations. We rather exploit the fact that the covariance matrix of the monthly data can be consistently estimated despite the missing values at the end of the sample. To estimate the covariance, we simply compute pairwise covariances over the periods both series are available. Thus, the EM algorithm is only used to interpolate the missing values and estimate the factors by the fixed weights times the data, which partly consists of estimated observations. We adopt this simplification to prevent convergence problems and to speed up the convergence process. As a stopping rule, we assume that convergence is achieved if the change in the average sum of squares of the idiosyncratic components is smaller than 10^{-5}.

Concerning the specifications of MIDAS, we use a large variety of initial parameter specifications, and compute the residual sum of squares (RSS). The parameter set with the smallest RSS then serves as the initial parameter set for NLS estimation. The parameters of the exponential lag function are restricted to $\theta_1 < 2/5$ and $\theta_1 < 0$, in line with Ghysels et al. (2007). The maximum number of lags chosen for MIDAS is $K = 12$ months.

3.2 Empirical results: A comparison of factor estimation methods for ragged-edge data

Now- and forecast results for the different combinations of MIDAS projections and factor estimation methods can be found in table 1. The table is divided into four parts. For each of the four MIDAS projections, we can compare the different factor estimation methods. The table shows relative MSEs to GDP variance and rankings based on those relative MSEs, where models with the smallest MSE rank first. The now- and forecast horizons are shown for monthly horizons $h_m = 1, \ldots, 9$, where horizons one to three belong to the nowcast. Horizon $h_m = 1$ is a nowcast made in the third month of the respective quarter, whereas horizon $h_m = 2$ is the nowcast made in the second month of the current quarter. Thus, similar to standard forecast comparisons, increasing horizons correspond to less information available for now- and forecasting, and we expect an increasing MSE for increasing horizons h_m.

The projections from the factor models have information content for the nowcast, as the MSEs of virtually all combinations of factor estimation methods and projection methods yield MSEs smaller than one, see table 1. For the one-quarter ahead forecast, we find borderline results. Comparing the factor estimation methods at horizons four to six, the results are not clear cut, where some relative MSEs are larger than one for some horizons and smaller for others. For two quarters ahead, the relative MSEs are for all factor models larger than one, thus rendering all factor models at hand uninformative for this horizon. This indicates that the methods employed here can be regarded as suited for short-term now- and forecasting only.

The differences between the factor estimation methods are relatively small overall. In the rankings of nowcast performance, EM-PCA factors do best in many cases in terms of ranking. However, for $h_m = 1$ and using the MIDAS-U and MIDAS-smooth projections together with factors VA-DPCA and KFS-PCA, respectively, do better than EM-PCA. Across projection methods, there are no systematic differences in nowcasting performance between factor estimation by VA-

Table 1: Comparison of nowcast and forecast results for different factor estimation methods for $r = 1$, MSE relative to GDP variance and ranking

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1.a. MIDAS-basic	VA-DPCA	0.71	1.01	1.06	0.94	1.18	1.05	1.16	1.24	1.30
	EM-PCA	0.62	0.69	0.78	1.07	0.99	1.01	1.30	1.09	1.05
	KFS-PCA	0.79	0.91	0.87	1.16	1.17	1.06	1.23	1.13	1.20
1.b. Ranking	VA-DPCA	2	3	3	1	3	2	1	3	3
	EM-PCA	1	1	1	2	1	1	3	1	1
	KFS-PCA	3	2	2	3	2	3	2	2	2
2.a. MIDAS-U	VA-DPCA	0.90	1.05	1.02	1.04	1.15	1.11	1.19	1.13	1.17
	EM-PCA	0.92	0.65	0.72	1.08	1.05	0.90	1.19	1.42	1.40
	KFS-PCA	0.89	0.90	0.81	0.97	1.03	1.02	1.31	1.49	1.36
2.b. Ranking	VA-DPCA	2	3	3	2	3	3	2	1	1
	EM-PCA	3	1	1	3	2	1	1	2	3
	KFS-PCA	1	2	2	1	1	2	3	3	2
3.a. MIDAS-smooth	VA-DPCA	0.69	0.92	0.87	0.95	1.10	1.20	1.18	1.12	1.19
	EM-PCA	0.70	0.73	0.84	0.94	0.95	1.00	1.05	1.09	1.13
	KFS-PCA	0.76	0.85	0.89	0.98	1.06	1.08	1.10	1.16	1.19
3.b. Ranking	VA-DPCA	1	3	2	2	3	3	3	2	3
	EM-PCA	2	1	1	1	1	1	1	1	1
	KFS-PCA	3	2	3	3	2	2	2	3	2
4.a. MIDAS-U0	VA-DPCA	0.71	0.86	0.89	0.90	1.05	0.98	1.05	1.09	1.12
	EM-PCA	0.58	0.65	0.72	0.92	0.93	0.79	1.10	1.10	1.05
	KFS-PCA	0.68	0.85	0.80	0.95	1.01	0.93	1.08	1.09	1.06
4.b. Ranking	VA-DPCA	3	3	3	1	3	3	1	2	3
	EM-PCA	1	1	1	2	1	1	3	3	1
	KFS-PCA	2	2	2	3	2	2	2	1	2

Note: The variance of GDP in the evaluation sample is 0.246. In the rankings, models with smallest MSE rank first. The model abbreviations are: VA-DPCA refers to the vertical realignment and dynamic PCA used in Altissimo et al. (2006), EM-PCA is the EM algorithm together with PCA as in Stock and Watson (2002), and KFS-PCA is the Kalman smoother of state-space factors according to Doz et al. (2006). The projection MIDAS-basic is the projection from Ghysels and Valkanov (2006), MIDAS-U is unrestricted MIDAS without exponential lag polynomial and lag specification using BIC, from Marcellino and Schumacher (2007). MIDAS-smooth is the projection as employed in Altissimo et al. (2006), and MIDAS-U0 is the MIDAS projection with unrestricted lag polynomials of order zero.

DPCA and KFS-PCA, as the relative MSE rankings change depending on the now- and forecast horizon.

Note that in table 1 increasing the nowcast or forecast horizon month by month not always leads to an increase of relative MSE, although this happens in most of the cases. This can be observed across all models under comparison. Thus, as new monthly information becomes available, the methods employed here cannot always improve the now- and forecasts with this information. This could be due to the relatively short sample under consideration, that induces high sampling uncertainty of the estimates and nowcasts.

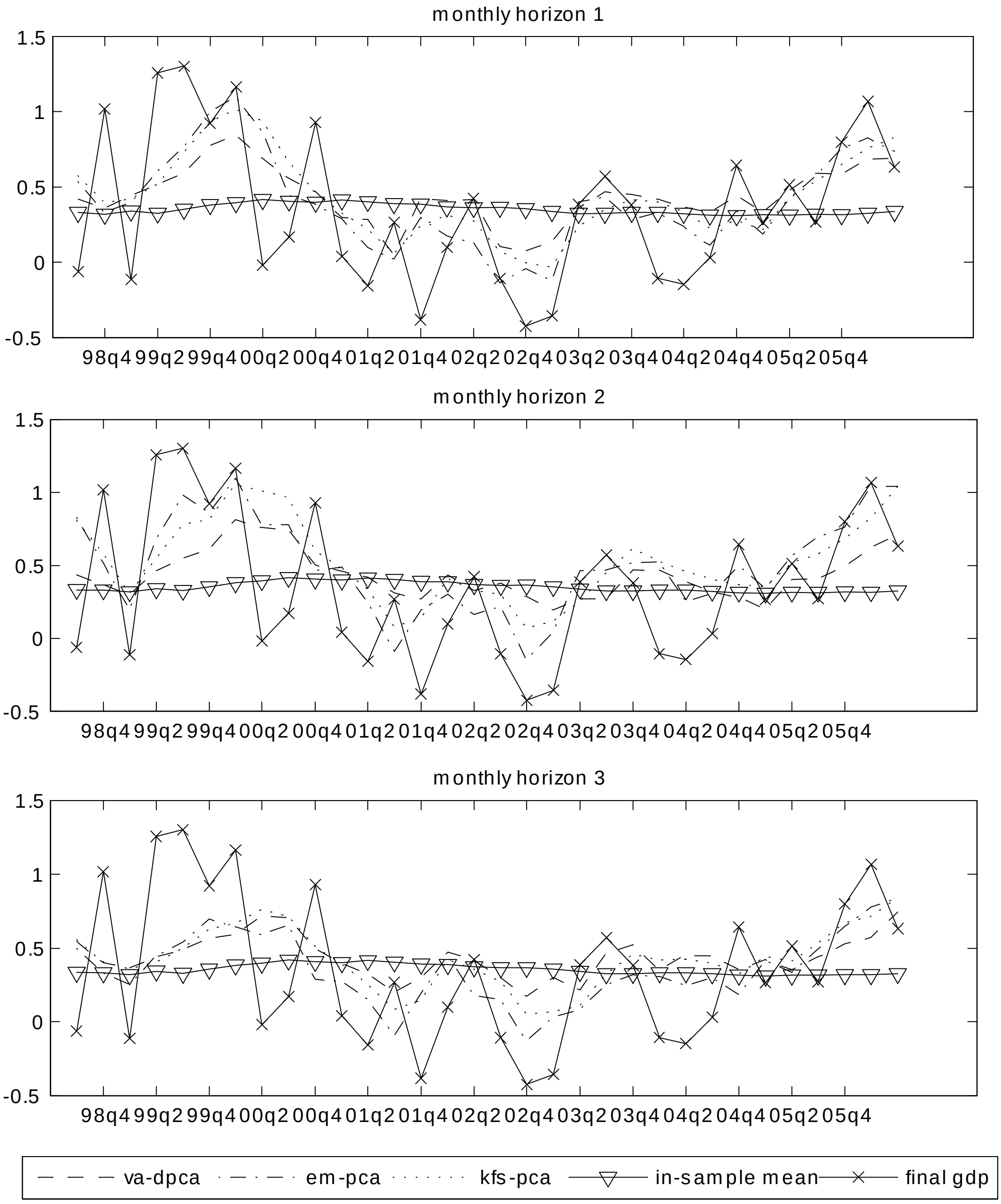
The relative comparison of the factor estimation methods was based on the MSE as a performance measure so far. However, as the MSE averages over observations in the evaluation period, this statistic can be dominated by differences in performance in only a few periods. Therefore, we additionally investigate the factor nowcasts over recursions. In figure 1, the time series of nowcasts for $h_m = 1, 2, 3$ are shown together with GDP observations and the in-sample mean as a benchmark nowcast for different factor estimation methods. Concerning the type of projection, figure 1 includes results for MIDAS-U0 only. As the results are very similar for the other types of MIDAS projections, we leave them out of this comparison here. The same holds for the forecast horizons $h_m \geq 3$. For comparative purposes, we include the in-sample mean of GDP as a naive nowcast into the figures. The results in figure 1 show that the three factor models perform clearly better than the simple benchmark. However, the erratic movements of GDP growth at the beginning of the sample, for example in $2000Q2$ and $2000Q3$, are not predicted well by all three factor models. Increasing the nowcast horizon from $h_m = 1$ to $h_m = 3$ shows the decline in variance of the nowcasts and, thus, a decline in nowcast ability. A common finding of the figures is the high correlation between the forecasts of the three factor models, as periods of good performance and periods of bad performance are similar. Therefore, in line with the similar MSE findings above, we find no clear indications of dramatic differences between the nowcast accuracy of the three factor models over time.

3.3 Empirical results: A comparison of MIDAS projections

Below, we discuss the different types of MIDAS projections. The nowcast results can be found in table 2. The table contains three groups for each of the factor estimation method. For each factor estimation method, we will compare the different MIDAS projections.

In table 2, a general finding is that the differences between the MIDAS approaches are not big as all approaches lead to nowcasts that have information content for current GDP, and only a few combinations of factor estimation and MIDAS projection also have predictive ability for the next quarter. Comparing the methods, we see that the difference between MIDAS-basic based on exponential lags and MIDAS-U is not clear-cut, as none of them outperforms the other across

Figure 1: Nowcasts with MIDAS-U0 and different factor estimation methods for horizon $h_m = 1, 2, 3$ and GDP observations, quarter on quarter growth, number of factors $r = 1$ and $q = 1$



Note: The figure shows nowcasts for the different factor estimation methods and the in-sample mean as a benchmark. For the model descriptions and abbreviations, see table 1.

Table 2: Comparison of nowcast and forecast results from different MIDAS projections for $r = 1$, MSE relative to GDP variance and ranking

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1.a. VA-DPCA	MIDAS-basic	0.71	1.01	1.06	0.94	1.18	1.05	1.16	1.24	1.30
	MIDAS-U	0.90	1.05	1.02	1.04	1.15	1.11	1.19	1.13	1.17
	MIDAS-smooth	0.69	0.92	0.87	0.95	1.10	1.20	1.18	1.12	1.19
	MIDAS-U0	0.71	0.86	0.89	0.90	1.05	0.98	1.05	1.09	1.12
1.b. Ranking	MIDAS-basic	3	3	4	2	4	2	2	4	4
	MIDAS-U	4	4	3	4	3	3	4	3	2
	MIDAS-smooth	1	2	1	3	2	4	3	2	3
	MIDAS-U0	2	1	2	1	1	1	1	1	1
2.a. EM-PCA	MIDAS-basic	0.62	0.69	0.78	1.07	0.99	1.01	1.30	1.09	1.05
	MIDAS-U	0.92	0.65	0.72	1.08	1.05	0.90	1.19	1.42	1.40
	MIDAS-smooth	0.70	0.73	0.84	0.94	0.95	1.00	1.05	1.09	1.13
	MIDAS-U0	0.58	0.65	0.72	0.92	0.93	0.79	1.10	1.10	1.05
2.b. Ranking	MIDAS-basic	2	3	3	3	3	4	4	2	1
	MIDAS-U	4	1	1	4	4	2	3	4	4
	MIDAS-smooth	3	4	4	2	2	3	1	1	3
	MIDAS-U0	1	2	2	1	1	1	2	3	2
3.a. KFS-PCA	MIDAS-basic	0.79	0.91	0.87	1.16	1.17	1.06	1.23	1.13	1.20
	MIDAS-U	0.89	0.90	0.81	0.97	1.03	1.02	1.31	1.49	1.36
	MIDAS-smooth	0.76	0.85	0.89	0.98	1.06	1.08	1.10	1.16	1.19
	MIDAS-U0	0.68	0.85	0.80	0.95	1.01	0.93	1.08	1.09	1.06
3.b. Ranking	MIDAS-basic	3	4	3	4	4	3	3	2	3
	MIDAS-U	4	3	2	2	2	2	4	4	4
	MIDAS-smooth	2	2	4	3	3	4	2	3	2
	MIDAS-U0	1	1	1	1	1	1	1	1	1

Note: For model abbreviations, see table 1.

all factor estimation methods and horizons. The most simple MIDAS projections without lags of the factors, MIDAS-U0 and MIDAS-smooth, provide often better nowcasts than MIDAS based on exponential lag functions, MIDAS-basic or MIDAS-U. MIDAS-smooth can outperform MIDAS-U0 only for factors obtained by VA-DPCA for a few horizons. However, based on EM-PCA and KFS-PCA factors, the projection MIDAS-U0 outperforms MIDAS-smooth at all horizons. Thus, the simplest projection method MIDAS-U0 seems to work best overall, as it ranks first or second in most of the cases.

As an extension to the basic MIDAS approach, Clements and Galvão (2007) consider autoregressive dynamics in the MIDAS approach. In particular, they propose the model

$$y_{t_m+h_m} = \beta_0 + \lambda y_{t_m} + \sum_{i=1}^{r} \beta_{1,i} b_i(L_m, \boldsymbol{\theta}_i)(1 - \lambda L_m^3)\widehat{f}_{i,t_m}^{(3)} + \varepsilon_{t_m+h_m}. \tag{17}$$

The autoregressive coefficient λ is not estimated unrestrictedly to rule out discontinuities of the impulse response function of $\widehat{\mathbf{F}}_{t_m}^{(3)}$ on $y_{t_m+h_m}$, see the discussion in Ghysels et al. (2007), pp. 60. The restriction on the coefficients is a common-factor restriction to ensure a smooth impulse response function, see Clements and Galvão (2007). The AR coefficient λ can be estimated together with the other coefficients by NLS. As an AR model is often supposed to be an appropriate benchmark specification for GDP, the extension of MIDAS might give additional insights in which direction the other MIDAS approaches considered so far might be improved.

In table 3 below, we will denote this variant as 'MIDAS-AR'. It is compared with the MIDAS-basic without AR terms. The results in table 3 show that considering AR terms doesn't improve the now- and forecast performance systematically. For different horizons and different factor estimation methods, the ranking between MIDAS-AR and MIDAS-basic changes. MIDAS-AR is not generally better than MIDAS-basic, which might also indicate problems with estimating autoregressive dynamics in German GDP. Note that we also tried to augment the unrestricted MIDAS with AR terms. However, also this experiment didn't lead to clear-cut improvements in forecast performance as well.

3.4 Empirical results: A comparison of monthly factor nowcast models with quarterly factor models

We now investigate the relative advantages of the nowcast factor models with earlier factor approaches in the literature. A widely followed way in the previous literature on factor forecasting is time aggregation. To obtain a balanced sample of data, one can simply aggregate the monthly data to quarterly data and ignore the most recent observations of high-frequency indicators. Then, the standard techniques of factor forecasting with single-frequency data can be employed. Note that previously most of the studies for forecasting of German GDP were based on quarterly, partly

Table 3: MIDAS-AR versus MIDAS-basic, comparison of relative MSE for $r = 1$

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1.a. VA-DPCA	MIDAS-AR	0.76	0.87	0.91	1.02	1.16	1.05	1.20	1.25	1.29
	MIDAS-basic	0.71	1.01	1.06	0.94	1.18	1.05	1.16	1.24	1.30
1.b. Ranking	MIDAS-AR	2	1	1	2	1	2	2	2	1
	MIDAS-basic	1	2	2	1	2	1	1	1	2
2.a. EM-PCA	MIDAS-AR	0.66	0.63	0.74	1.12	1.13	0.96	1.24	1.10	1.35
	MIDAS-basic	0.62	0.69	0.78	1.07	0.99	1.01	1.30	1.09	1.05
2.b. Ranking	MIDAS-AR	2	1	1	2	2	1	1	2	2
	MIDAS-basic	1	2	2	1	1	2	2	1	1
3.a. KFS-PCA	MIDAS-AR	0.88	0.93	0.84	1.08	1.18	1.13	1.28	1.16	1.26
	MIDAS-basic	0.79	0.91	0.87	1.16	1.17	1.06	1.23	1.13	1.20
3.b. Ranking	MIDAS-AR	2	2	1	1	2	2	2	2	2
	MIDAS-basic	1	1	2	2	1	1	1	1	1

Note: The projection MIDAS-AR contains one autoregressive term as in Clements and Galvão (2007), MIDAS basic is wihout AR terms. For factor model abbreviations, see table 1.

time-aggregated data, see e.g. Schumacher (2007). As quarterly data is widely used in the empirical literature for GDP forecasting, we will also compare the mixed-frequency nowcast models to the quarterly factor models.

In particular, we employ the standard model for factor forecasting following Stock and Watson (2002). The forecast equation is essentially a quarterly factor-augmented AR model according to

$$y_{t_q+h_q} = \beta_0 + \lambda(L_q)y_{t_q} + \mathbf{E}(L_q)\widehat{\mathbf{F}}^Q_{t_q} + \varepsilon_{t_q+h_q}, \tag{18}$$

where $\mathbf{E}(L_q) = \sum_{p=0}^{P} \mathbf{E}_p L_q^p$ is an unrestricted lag polynomial of order P, and L_q is the quarterly lag operator now. $\lambda(L_q)$ is now a lag polynomial of order R for autoregressive terms. The factors $\widehat{\mathbf{F}}^Q_{t_q}$ are estimated by PCA, which is applied to the quarterly indicators. These time series indicators are the same as for the nowcast models as discussed above, but aggregated over time to quarterly frequency. Note that model (18) with static factors $\widehat{\mathbf{F}}^Q_{t_q}$ works quite well for single-frequency data compared with dynamic factor estimates, see Boivin and Ng (2006), D'Agostino and Giannone (2006), as well as Schumacher (2007) for German GDP. Thus, it might serve as an interesting alternative to the nowcast models.

As a benchmark for the factor nowcast models, we employ a univariate quarterly autoregressive

(AR) model for GDP, specified using the BIC with a maximum lag order of three quarters. It turns out that in almost all of the recursions, only one lag is chosen. Furthermore, we present the in-sample mean of GDP as an additional benchmark. In the recent forecasting literature, this benchmark has turned out to be a strong competitor to more sophisticated approaches, see e.g. De Mol et al. (2006).

Table 4 contains results for the nowcasts using quarterly factor models as well as the simple benchmarks. As representatives of the nowcast models, we present results based on MIDAS-U0 and the three different ragged-edge factor estimation methods. In the empirical nowcast comparison,

Table 4: Comparison of mixed-frequency nowcast models with MIDAS-U0 and quarterly factor and benchmark models, MSE relative to GDP variance and ranking

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1. MIDAS-U0	VA-DPCA	0.71	0.86	0.89	0.90	1.05	0.98	1.05	1.09	1.12
	EM-PCA	0.58	0.65	0.72	0.92	0.93	0.79	1.10	1.10	1.05
	KFS-PCA	0.68	0.85	0.80	0.95	1.01	0.93	1.08	1.09	1.06
2. Quarterly factor model	PCA, $P = R = 0$, $r = 1$	0.98	1.05	1.05	1.05	1.16	1.16	1.16	1.14	1.14
	PCA, $P = R = 0$, $r = 2$	1.03	0.94	0.94	0.94	1.31	1.31	1.31	1.25	1.25
	PCA-BIC, $r = 1$	0.91	1.02	1.02	1.02	1.13	1.13	1.13	1.08	1.08
	PCA-BIC, $r = 2$	0.94	0.89	0.89	0.89	1.23	1.23	1.23	1.23	1.23
	PCA-BIC	0.99	1.06	1.06	1.06	1.13	1.13	1.13	1.08	1.08
3. Benchmarks	AR	1.02	1.17	1.17	1.17	1.08	1.08	1.08	1.08	1.08
	in-sample mean	1.03	1.04	1.04	1.04	1.05	1.05	1.05	1.06	1.06

Note: The quarterly factor model contains factors estimated from quarterly, time-aggregated data using PCA. The forecast equation The first two specifications are based on a fixed number of factors and a fixed number of lags, whereas the third and fourth are based on a fixed number of factors and the number of lags is chosen by BIC. PCA-BIC selects the number of factors as well as the lag orders using BIC as in Stock and Watson (2002). Concerning the nowcast models, the abbreviations are explained in table 1.

the simple benchmarks do not perform well, as can be seen from the bottom rows of table 4. Both the AR model and the in-sample mean have relative MSEs larger than one. Note that, whereas the nowcast factor models employ monthly information which is updated every month and, thus, can lead to changes in now- and forecast MSEs, the benchmark models and the quarterly factor models change only every third month (when a new observations of GDP is available), implying a constant MSE for three months.

The quarterly factor model performs better than the naive benchmarks, and has some information content for GDP for horizons up to three months. For longer horizons, there is almost no information content in the forecasts. Compared with the monthly nowcast models, the quarterly factor model is generally outperformed for the nowcast for $h_m \leq 3$. In many cases, this also holds for the one-quarter ahead forecast, although the differences are smaller at these horizons. Thus, according to these results, taking into account ragged-edge information as in the nowcast models with monthly indicators can improve the current estimate of GDP. As the use of time-aggregated data implies a loss of information at the end of the sample, the results imply that the nowcast methods employed here can to some extent exploit this information. In summary, we can confirm that in general it is advisable to employ the ragged-edge data together with the different factor estimation techniques for nowcasting.

3.5 Empirical results: Static versus dynamic factors

Following the discussion in Boivin and Ng (2005), there is some disagreement in the literature concerning the appropriate factor estimation method to be employed for forecasting. In particular, it is unclear whether DPCA or PCA are favourable for predictive purposes. In general, there is no consensus as to the appropriate estimation method, see also the discussion in Schneider and Spitzer (2004), Den Reijer (2005), D'Agostino and Giannone (2006), and again Boivin and Ng (2005) for different datasets. In a dataset for the German economy with balanced recursive samples, dynamic PCA does not generally work better, and the differences between the methods are small, see Schumacher (2007).

Against the background of this discussion, we will address this issue also in the present context. In our applications above, DPCA was employed to estimate the factors in combination with vertical realignment of the data. To compare the sensitivity of the results, we compare the existing results using VA-DPCA with static PCA and vertical realignment of the data, denoted as VA-PCA below. Table 5 shows relative MSEs to GDP variance for the different factor estimates and different projection techniques. The results show that the information content of the now- and forecasts does hardly change if the factors are estimated by PCA instead of DPCA. MSEs relative to GDP variance are in most of the cases above or below one for both factor estimators. The bottom part of the table shows another relative MSE defined as the MSE obtained from using DPCA factors divided by the MSE obtained from using static PCA factors for forecasting. The results show no systematic advantages over the horizons between the two methods. Thus, the way the factors are estimated seems to be of limited importance in this application.

3.6 Empirical results: Integrated state-space model approach versus two-step nowcasting

The results obtained so far are entirely based on a two-step procedure: The factors are estimated firstly, and then forecasting is carried out using the MIDAS approaches. However, among the

Table 5: Static PCA versus dynamic PCA nowcasts for $r = 1$, MSE relative to GDP variance in part 1. to 5., part 6 DPCA MSE divided by PCA MSE

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1. MIDAS-basic	VA-DPCA	0.71	1.01	1.06	0.94	1.18	1.05	1.16	1.24	1.30
	VA-PCA	0.69	1.05	1.02	0.99	1.17	1.04	1.07	1.24	1.35
2. MIDAS-U	VA-DPCA	0.90	1.05	1.02	1.04	1.15	1.11	1.19	1.13	1.17
	VA-PCA	0.76	1.14	1.00	1.00	1.12	1.08	1.15	1.12	1.18
3. MIDAS-smooth	VA-DPCA	0.69	0.92	0.87	0.95	1.10	1.20	1.18	1.12	1.19
	VA-PCA	0.70	0.98	0.88	0.93	1.07	1.12	1.14	1.07	1.17
4. MIDAS-U0	VA-DPCA	0.71	0.86	0.89	0.90	1.05	0.98	1.05	1.09	1.12
	VA-PCA	0.69	0.93	0.93	0.85	1.08	0.91	1.04	1.07	1.13
6. Relative MSE:	MIDAS-basic	1.04	0.96	1.04	0.96	1.01	1.01	1.08	1.01	0.96
DPCA/PCA	MIDAS-U	1.19	0.92	1.02	1.04	1.02	1.03	1.03	1.01	0.99
	MIDAS-smooth	0.99	0.93	0.98	1.02	1.02	1.08	1.04	1.05	1.02
	MIDAS-U0	1.03	0.93	0.96	1.06	0.96	1.08	1.00	1.02	0.99

Note: Parts one to five show relative MSEs to variance of GDP. Part six shows another relative MSE defined as the MSE of the VA-DPCA factor model divided by the MSE of the model using static factors, denoted as VA-PCA. For model and projection abbreviations, see table 1.

models, the state-space approach allows in general for joint estimation of the factors and nowcasting GDP, see Giannone et al. (2005). For the Euro area, Banbura and Rünstler (2007) propose to augment the state-space model by a simple static relationship between monthly GDP and the factors. This follows the seminal work by Mariano and Murasawa (2003), where combining monthly and quarterly data in a small factor state-space model has been introduced.

In particular, Banbura and Rünstler (2007) augment the state-space system above, see equations (5) and (6), with further relationships that interpolate GDP and relate monthly GDP to the monthly factors. All in all, they add three equations, see Banbura and Rünstler (2007), p. 5: Equation 1) $y_{t_q} = \widetilde{y}_{t_q} + \varepsilon_{t_q}$, with ε_{t_q} as a measurement error, which is normally distributed with mean zero and variance Σ_ε; 2) an equation for time aggregation $\widetilde{y}_{t_q} = \widetilde{y}_{t_m} = (\frac{1}{3} + \frac{2}{3}L_m + L_m^2 + \frac{2}{3}L_m^3 + \frac{1}{3}L_m^4)y_{t_m}^m$ for $t_m = 3, 6, \ldots, T_m$, and 3) the static factor representation at the monthly frequency $y_{t_m}^m = \mathbf{\Lambda}_y \mathbf{F}_{t_m}$. Equations 2) and 3) add to the vector state equation, whereas 1) adds to the vector observation equation of the state space model. In line with the estimation procedure for the factor-only state-space model (5) and (6) above, Banbura and Rünstler (2007) estimate the coefficients $\mathbf{\Lambda}_y$, Σ_ε outside the state-space model by estimating a reduced form of 1) to 3), which is a regression model for quarterly GDP dependent on time-aggregated quarterly factors. They plug the resulting estimates of $\mathbf{\Lambda}_y$ and Σ_ε in the state-space model for Kalman filtering and smoothing, which now also provides the now- and forecasts for GDP, as y_{t_q} is part of the observation vector in this integrated approach.

The key difference between the two-step factor-estimation MIDAS approach chosen in the applications above and the ones followed by Banbura and Rünstler (2007) and Mariano and Murasawa (2003) is that MIDAS directly relates time series of different frequencies, whereas the state-space approaches allow for specifying relationships consistently at the higher frequency. Furthermore, MIDAS is a direct forecast device, whereas the Kalman smoother is based on a VAR model that yields iterative forecasts in the terminology of Marcellino et al. (2006). This approach is fully integrated as it interpolates missing values of the indicators, estimates factors and yields nowcasts of GDP in one coherent framework. To check whether this strategy can improve over the two-step approach followed here so far in terms of now- and forecasting, we also provide nowcast results for the model proposed by Banbura and Rünstler (2007). Table 6 shows relative MSEs to GDP variance and rankings for the different state-space model now- and forecasts. In the table, 'KFS-PCA full' denotes the fully-integrated approach, whereas all the other forecasts are based on the two-step procedure, where the Kalman smoother is used to estimate the monthly factors only. Note that the coefficients of the state-space model are reestimated for each recursion in the exercise. Therefore, factors estimates can change due to parameter changes as well as the addition of new information at the end of the sample. The results show that the integrated approach also does well in now- and forecasting. It performs better than the two-step MIDAS-basic and MIDAS-U projection, and very similar to the simple MIDAS-U0 projection. For horizons two,

Table 6: Two-step KFS-PCA vs fully integrated now- and forecast results from the state-space model for $r = 1$, MSE relative to GDP variance and ranking

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1.a. Relative MSE	KFS-PCA full	0.70	0.81	0.84	0.88	1.00	0.95	1.10	1.12	1.09
	MIDAS-basic	0.79	0.91	0.87	1.16	1.17	1.06	1.23	1.13	1.20
	MIDAS-U	0.89	0.90	0.81	0.97	1.03	1.02	1.31	1.49	1.36
	MIDAS-smooth	0.76	0.85	0.89	0.98	1.06	1.08	1.10	1.16	1.19
	MIDAS-U0	0.68	0.85	0.80	0.95	1.01	0.93	1.08	1.09	1.06
1.b. Ranking	KFS-PCA full	2	1	3	1	1	2	3	2	2
	MIDAS-basic	4	5	4	5	5	4	4	3	4
	MIDAS-U	5	4	2	3	3	3	5	5	5
	MIDAS-smooth	3	3	5	4	4	5	2	4	3
	MIDAS-U0	1	2	1	2	2	1	1	1	1

Note: For model abbreviations, see table 1.

four and five, it performs best among all the different approaches. For horizons, one and three, the MIDAS-U0 performs best.

The similar performance of the fully integrated state-space model to the very simple MIDAS projections confirms the previous findings that simple and very parsimonious projection models seem to work better than more complicated models. Note that the equation $y^m_{t_m} = \mathbf{\Lambda}_y \mathbf{F}_{t_m}$ in the state-space model above, that relates monthly GDP and the factors, is very parsimonious and does not contain lags of the factors as is the case of the MIDAS-U0 forecast. Whether the approach is integrated within one coherent state-space model or split into two steps is, however, of second order importance according to our findings. Therefore, we do not seem to loose much if we rely on the two-step procedure, which allows us to compare the different factor estimation methods.

4 Conclusions

The nowcasting perspective followed in this paper takes into account the publication lags of statistical data that decision makers face in their everyday business of assessing the current state of the economy. Due to the publication delay of GDP, the necessity of nowcasting as a projection of current quarter GDP directly emerges, and specific solutions are needed that can employ information from many business cycle indicators, that are also subject to publication lags and thus lead to the so-called 'ragged-edge' of the data.

The factor models and projection methods discussed here can tackle these nowcasting issues.

Based on the two-step procedure often followed in the recent factor-forecasting literature, we differentiate between the factor-estimation step and the factor-forecasting step. When estimating the factors, we place special emphasis on missing values at the end of the sample due to statistical publication lags. Regarding the factor-forecasting step, we introduce the Factor-MIDAS approach as a simple tool for direct now- and forecasting in a mixed-frequency context.

The different nowcast approaches are applied to a German post-unification dataset, and compared with respect to their nowcasting performance of German GDP growth. The results indicate that all the nowcast models can improve over quarterly factor forecasts based on time-aggregated data. Thus, taking into account the ragged-edge of the data and exploiting most recent observations pays off to some extent for nowcasting.

Concerning the differences between the MIDAS projection methods, the results indicate that MIDAS with exponential distributed lag functions performs similarly to MIDAS with unrestricted lag polynomials. The best performing projection is in many cases a very simple MIDAS without a distributed lag structure and only up to one lag of the factors. Autoregressive dynamics also play only a minor role in the projections.

The choice of the factor estimation techniques that can tackle missing values has no substantial impact on the nowcast performance. The EM algorithm together with static PCA as in Stock and Watson (2002), vertical realignment together with dynamic PCA as in Altissimo et al. (2002), as well as factors estimated using a large state-space model with QML as in Doz et al. (2006) all provide informative nowcasts and to a lesser extent informative forecasts one quarter ahead. Compared with respect to their performance over time, we observe that the forecasts based on the three factor estimation methods are highly correlated. There are also no systematic differences between static and dynamic PCA for nowcasting. Interestingly, factor-forecast applications with single-frequency data, see e.g. D'Agostino and Giannone (2006) and Schumacher (2007), have recently obtained similar findings. A final results from the application here is that choosing an integrated state-space model rather than the two-step procedure followed here cannot improve the nowcast performance.

Although there are clear now- and forecast gains from the application of the factor models discussed here at short horizons, the same does not hold for the longer forecast horizons of up to two quarters. At these horizons, the forecasts of all the factor models are hardly uninformative. Therefore, the methods employed here can only be regarded as short-term forecasting devices, and there is room for improvements of the methods for longer horizons. Note, however, that this is a problem that can often be observed in the recent literature. Related to the debate on the 'Great Moderation', there is evidence of a decline in forecastability of real and nominal variables for many sophisticated forecast procedures, see D'Agostino et al. (2006) and Campbell (2007), for example.

References

[1] Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., Veronese, G. (2006), New Eurocoin: Tracking Economic Growth in Real Time, CEPR Working Paper 5633.

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

[3] Bai, J., Ng, S. (2002), Determining the Number of Factors in Approximate Factor Models, Econometrica 70, 191-221.

[4] Bai, J., Ng, S. (2006), Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions, Econometrica, 74, 1133-1150.

[5] Bai, J., Ng, S. (2007), Determining the Number of Primitive Shocks in Factor Models, Journal of Business and Economic Statistics 25, 52-60.

[6] Banbura, M., Rünstler, G. (2007), A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP, ECB Working Paper 751.

[7] Banerjee, A., Marcellino, M. (2006), Are there any reliable leading indicators for US inflation and GDP growth?, International Journal of Forecasting 22, 137-151.

[8] Banerjee, A., Marcellino, M., Masten, I. (2005), Leading Indicators for Euro-area Inflation and GDP Growth, Oxford Bulletin of Economics and Statistics, 67, 785-813.

[9] Banerjee, A., Marcellino, M., Masten, I. (2007), Forecasting macroeconomic variables using diffusion indexes in short samples with structural change, in Rapach, D., Wohar, M. (eds.), Forecasting in the Presence of Structural Breaks and Model Uncertainty, Elsevier, forthcoming.

[10] Bernanke, B., Boivin, J. (2003), Monetary Policy in a Data-Rich Environment, Journal of Monetary Economics 50, 525-546.

[11] Boivin, J., Ng, S. (2005), Understanding and Comparing Factor-Based Forecasts, International Journal of Central Banking 1, 117-151.

[12] 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.

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

[14] Clements, M., Galvão, A. (2007), Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US output growth, Queen Mary University of London Working Paper 616.

[15] D'Agostino, A., Giannone, D., Surico, P. (2006), (Un)Predictability and macroeconomic stability, ECB Working Paper 605.

[16] D'Agostino, A., Giannone, D. (2006), Comparing alternative predictors based on large-panel factor models, Working Paper 680.

[17] De Mol, C., Giannone, D., Reichlin, R. (2006), Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?, ECB Working Paper 700.

[18] Den Reijer, A. (2005), Forecasting Dutch GDP using large scale factor models, DNB Working Paper 28.

[19] Doz, C., Giannone, D., Reichlin, L. (2006), A quasi maximum likelihood approach for large approximate dynamic factor models, ECB Working Paper 674.

[20] Eickmeier, S., Ziegler, C. (2007), How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach, Journal of Forecasting, forthcoming.

[21] Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2003), Do financial variables help forecasting inflation and real activity in the euro area? Journal of Monetary Economics 50, 1243 - 1255.

[22] Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2005), The Generalized Dynamic Factor Model: One-sided estimation and forecasting, Journal of the American Statistical Association 100, 830-840.

[23] Ghysels, G., Valkanov, R. (2006), Linear Time Series Processes with Mixed Data Sampling and MIDAS Regression Models, mimeo.

[24] Ghysels, G., Sinko, A., Valkanov, R. (2007), MIDAS Regressions: Further Results and New Directions, Econometric Reviews 26, 53-90.

[25] 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.

[26] Kapetanios, G., Labhard, V., Price, S. (2007), Forecast combination and the Bank of England's suite of statistical forecasting models, Bank of England Working Paper no. 323.

[27] Lütkepohl, H. and Brüggemann, R. (2006), A small monetary system for the euro area based on German data, Journal of Applied Econometrics, 21, 683-702.

[28] Marcellino, M., Kapetanios, G. (2006), A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions, CEPR WP 5620.

[29] Marcellino, M., Schumacher, C. (2007), MIDAS and Dynamic Linear Models, mimeo.

[30] Marcellino, M., Stock, J., Watson, M. (2005), Macroeconomic forecasting in the Euro area: country specific versus Euro wide information, European Economic Review 47, 1-18.

[31] 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.

[32] 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.

[33] Schneider, M., Spitzer, M. (2004), Forecasting Austrian GDP using the generalized dynamic factor model, OeNB Working Paper 89.

[34] Schumacher, C. (2007), Forecasting German GDP using alternative factor models based on large datasets, Journal of Forecasting 26, 271-302.

[35] Schumacher, C., Breitung, J. (2006), Real-time forecasting of GDP based on a large factor model with monthly and quarterly data, Deutsche Bundesbank Discussion Paper, Series 1: Economic Studies, No. 33/06.

[36] Stock, J., Watson, M. (2002), Macroeconomic Forecasting Using Diffusion Indexes, Journal of Business & Economic Statistics 20, 147-162.

[37] Stock, J., Watson, M. (2006), Forecasting with Many Predictors, in: Elliot, G., Granger, C., Timmermann, A. (eds.), Handbook of Economic Forecasting, Vol 1, 515-554.

[38] Wallis, K. (1986), Forecasting with an econometric model: The 'ragged edge' problem, Journal of Forecasting 5, 1-13.

[39] Watson, M. (2003), Macroeconomic Forecasting Using Many Predictors, in: Dewatripont, M., Hansen, L., Turnovsky, S. (eds.), Advances in Economics and Econometrics, Theory and Applications, Eight World Congress of the Econometric Society, Vol 3, 87-115.

A Monthly dataset

This appendix describes the time series for the German economy used in the forecasting exercise. The whole data set for Germany contains 111 monthly time series over the sample period from $1992M1$ until $2006M11$. The time series cover broadly the following groups of data: prices, labour market data, financial data (interest rates, stock market indices), industry statistics, construction statistics, surveys and miscellaneous indicators.

The source of the time series is the Bundesbank database. The download date of the dataset is 6th December 2006. In this dataset, there are differing missing values at the end of the sample. For example, whereas financial time series are available up to $2006M11$, industrial time series like production, orders and so on are only available up to $2006M09$. This leads to a ragged-edge structure at the end of the sample, which serves as a template to replicate the ragged edges in past pseudo real-time periods as described in the main text.

Natural logarithms were taken for all time series except interest rates. Stationarity was obtained by appropriately differencing the time series. Most of the time series taken from the above source are already seasonally adjusted. Remaining time series with seasonal fluctuations were adjusted using Census-X12 prior to the forecast simulations. Extreme outlier correction was done using a modification of the procedure proposed by Watson (2003). Large outliers are defined as observations that differ from the sample median by more than six times the sample interquartile range (Watson, 2003, p. 93). The identified observation is set equal to the respective outside boundary of the interquartile.

A.1 Prices

producer price index
producer price index without energy
consumer price index
consumer price index without energy
export prices
import prices
oil price Brent GB

A.2 Labour market

unemployed
unemployment rate
employees and self-employed
employees, short-term

productivity, per employee
productivity, per hour
wages and salaries per employee
wages and salaries per hour
vacancies

A.3 Interest rates, stock market indices

money market rate, overnight deposits
money market rate, 1 month deposits
money market rate, 3 months deposits
bond yields on public and non-public long term bonds with average maturity from 1 to 2 years
bond yields on public and non-public long term bonds with average maturity from 5 to 6 years
bond yields on public and non-public long term bonds with average maturity from 9 to 10 years
yield spread: bond yields with maturity from 1 to 2 years minus 3 months money market rate
yield spread: bond yields with maturity from 5 to 6 years minus 3 months money market rate
yield spread: bond yields with maturity from 9 to 10 years minus 3 months money market rate
CDAX share price index
DAX German share index
REX German bond index
exchange rate US dollar/Deutsche Mark
indicator of the German economy's price competitiveness against 19 industrial countries based on consumer prices
monetary aggregate M1
monetary aggregate M2
monetary aggregate M3

A.4 Manufacturing turnover, production and received orders

production: intermediate goods industry
production: capital goods industry
production: durable and non-durable consumer goods industry
production: mechanical engineering
production: electrical engineering
production: vehicle engineering
export turnover: intermediate goods industry
domestic turnover: intermediate goods industry
export turnover: capital goods industry

domestic turnover: capital goods industry
export turnover: durable and non-durable consumer goods industry
domestic turnover: durable and non-durable consumer goods industry
export turnover: mechanical engineering
domestic turnover: mechanical engineering
export turnover: electrical engineering industry
domestic turnover: electrical engineering industry
export turnover: vehicle engineering industry
domestic turnover: vehicle engineering industry
orders received by the intermediate goods industry from the domestic market
orders received by the intermediate goods industry from abroad
orders received by the capital goods industry from the domestic market
orders received by the capital goods industry from abroad
orders received by the consumer goods industry from the domestic market
orders received by the consumer goods industry from abroad
orders received by the mechanical engineering industry from the domestic market
orders received by the mechanical engineering industry from abroad
orders received by the electrical engineering industry from the domestic market
orders received by the electrical engineering industry from abroad
orders received by the vehicle engineering industry from the domestic market
orders received by the vehicle engineering industry from abroad
industrial production

A.5 Construction

orders received by the construction sector: building construction
orders received by the construction sector: civil engineering
orders received by the construction sector: residential building
orders received by the construction sector: non-residential building construction
man-hours worked in building construction
man-hours worked in civil engineering
man-hours worked in residential building
man-hours worked in industrial building
man-hours worked in public building
turnover: building construction
turnover: civil engineering
turnover: residential building

turnover: industrial building

turnover: public building

production in the construction sector

A.6 Surveys

ifo surveys: business situation: capital goods producers

ifo surveys: business situation: producers durable consumer goods

ifo surveys: business situation: producers non-durable consumer goods

ifo surveys: business situation: retail trade

ifo surveys: business situation: wholesale trade

ifo surveys: business expectations for the next six months: producers capital goods

ifo surveys: business expectations for next six months: producers durable consumer goods

ifo surveys: business expectations for next six months: producers non-durable consumer goods

ifo surveys: business expectations for next six months: retail trade

ifo surveys: business expectations for next six months: wholesale trade

ifo surveys: stocks of finished goods: producers of capital goods

ifo surveys: stocks of finished goods: producers of durable consumer goods

ifo surveys: stocks of finished goods: producers of non-durable consumer goods

GfK consumer surveys: income expectations

GfK consumer surveys: business cycle expectations

GfK consumer surveys: propensity to consume: consumer climate

GfK consumer surveys: price expectations

ZEW financial market survey: business cycle expectations

A.7 Miscellaneous indicators

current account: exports

current account: imports

current account: services import

current account: services export

current account: transfers from abroad

current account: transfers to foreign countries

HWWA raw material price index

HWWA raw material price index without energy

HWWA raw material price index: industrial raw materials

HWWA raw material price index: energy industrial raw materials

new car registrations

new car registrations by private owners
retail sales turnover

B Nowcast results for different specifications of the factor models

This section presents nowcast and forecast results for different specifications of the factor models in terms of different numbers of static factors r and dynamic shocks q. Of course, estimation of the factors based on vertically realigned data and dynamic PCA (VA-DPCA) requires specification of both q and r, whereas the number of static factors r is the only auxiliary parameter for the factors that are estimated with the EM algorithm together with static PCA (EM-PCA). The factors estimated in the state-space model approach with the Kalman smoother (KFS-PCA) require specifying q and r. To check the sensitivity of the results with respect to the number of factors and shocks, we follow two specification schemes: Firstly, we compare fixed specifications, and, secondly, we employ information criteria for model specification.

Regarding the MIDAS projection, we report only results based on MIDAS-U0, which performs well compared with the other projections. Results are not shown for the other projections, as they lead to very similar conclusions.

B.1 Fixed specifications and information criteria

Concerning fixed specifications, we consider many combinations of the auxiliary parameters, as they can heavily influence the model's forecast performance, see Boivin an Ng (2005) for a discussion. In our application, we consider a maximum number of static factors of $r = 6$ and dynamic factors $q \leq 3$, and compute results for all possible combinations of the parameters. We considered also results for $3 < q \leq r$, but this didn't lead, in general, to improvements in nowcast performance, and we do not provide the results here.

Regarding the sensitivity analysis based on information criteria, we apply the ones proposed by Bai and Ng (2002, 2007). In particular, for the number of static factors, we adopt the IC_{p2} criterion of Bai and Ng (2002)

$$IC_{p2}(r) = \ln(V(r, \mathbf{F})) + r \left(\frac{N + T_m}{NT_m} \right) \ln(\min\{N, T_m\}). \tag{19}$$

The information criterion reflects the trade-off between goodness-of-fit and overfitting. The first term on the right-hand side shows the goodness-of-fit, which is given by the residual sum of squares

$$V(r, \mathbf{F}) = \frac{1}{NT_m} \sum_{i=1}^{N} \sum_{t_m=1}^{T_m} (x_{i,t_m} - \mathbf{\Lambda}_i \mathbf{F}_{t_m})^2 \,, \tag{20}$$

and depends on the estimates of the static factors and the number of factors. The residuals are given by $x_{i,t_m} - \mathbf{\Lambda}_i \mathbf{F}_{t_m}$, where $\mathbf{\Lambda}_i$ is a $(1 \times r)$ dimensional row vector of the parameter matrix $\mathbf{\Lambda}$ of the static model, see (1) in the main text. If the number of factors r is increased, the variance of the factors increases, too, and the sum of squared residuals decreases. Hence, the information criteria have to be minimised in order to determine the number of factors. The penalty for overfitting, which is the second term on the right-hand side behind r in (19), is an increasing function of the cross-section size N and time series length T_m. In empirical applications, one has to fix a maximum number of factors, say $r_{\max}$, and estimate the model for all number of factors $r = 1, \ldots, r_{\max}$. The optimal number of factors minimises IC_{p2}. In the forecast comparison, we set $r_{\max} = 6$. Note that T_m in IC_{p2} above is the time series sample size of the recursive subsample.

The number of dynamic shocks q for dynamic PCA estimation of the factors and the state-space model is determined by the information criterion proposed by Bai and Ng (2007). This criterion takes the estimated static factors as given, and estimates a VAR of lag order p on these factors, where p is determined by the Bayesian information criterion (BIC). Then, a spectral decomposition of the $(r \times r)$ residual covariance matrix $\widehat{\mathbf{\Gamma}}_u$ is computed, and $\widehat{c}_j$ is the j-th ordered eigenvalue, where $\widehat{c}_1 > \widehat{c}_2 \geq \ldots \geq \widehat{c}_r \geq 0$. Compute

$$\widehat{D}_k = \left(\frac{\widehat{c}_{k+1}}{\sum_{j=1}^{r} \widehat{c}_j} \right)^{1/2} \tag{21}$$

for $k = 1, \ldots, r-1$. Each $\widehat{D}_k$ is a measure of the marginal contribution of the respective eigenvalue, and under the assumption rank $(\widehat{\mathbf{\Gamma}}_u) = q$, $c_k = 0$ for $k > q$. Bai and Ng (2007) show that $\widehat{D}_k$ converges to zero for $k \geq q$. In applications, the set of admissible numbers of dynamic factors is chosen by a boundary according to $\mathcal{K} = \{k : \widehat{D}_k < m/\min[N^{2/5}, T^{2/5}]\}$. We use $m = 1.0$, following the Monte Carlo results in Bai and Ng (2007). Finally, the number of dynamic factors is given by $\widehat{q}^{BN} = \min\{k \in \mathcal{K}\}$.

In the tables below, the information criteria for r and q are applied recursively. Thus, the specifications can change over time in contrast to the specification with fixed numbers of factors and dynamic shocks.

B.2 Empirical results for the different factor models

Table 7 shows the nowcast results for different numbers of factors for the factors based on vertically realigned data and dynamic PCA (VA-DPCA) and MIDAS-U0 projection. In general, now- and forecasts with fewer factors r and a smaller number of shocks q are doing better than higher-dimensional model nowcasts for all the three MIDAS projections. For $r \geq 3$, most of the now- and forecasts are uninformative. Considering models whose performance is relatively stable across the horizons, models with $r = 1$, $q = 1$ and $r = 2$, $q = 1, 2$ do best in terms of ranking. Information criteria also do well in selecting models with high-ranking nowcast accuracy.

Table 8 shows the nowcast results for different numbers of static factors for the factors based on the EM algorithm and static PCA (EM-PCA). The results show, that in almost all of the cases, $r = 1$ is the best-performing specification. With a few exceptions, where $r = 2$ performs better, $r = 1$ has the most stable now- and forecast performance across horizons h_m. Information criteria tend to perform badly.

Table 9 shows the nowcast results for different numbers of factors for the state-space model approach with the Kalman smoother to estimate factors (KFS-PCA). The specification $r = 1$ and $q = 1$ is doing well for the nowcast. With $r = 2$ and $q = 1$ or $q = 2$, KFS-PCA also performs well, in some cases better than $r = 1$. Models specified using information criteria perform in most of the cases worse than models with only a few factors. Furthermore, the relative MSE to GDP variance is in almost all the cases larger than one, indicating uninformative now- and forecasts.

B.3 Summary of the comparison of specifications

The results of the sensitivity analysis lead to a clear-cut conclusion: If the number of factors is fixed larger than two, the now- and forecasts have in most of the cases no information content. Moreover, the information criteria select models, that have in most of the cases a poor performance, with the exception of the VA-DPCA factors. All the different factor models perform best with $r = 1$ or $r = 2$. As the results do not differ substantially across these specifications, in the main text we concentrate on the case $r = 1$ and $q = 1$.

Table 7: Nowcast and forecast results for VA-DPCA factors and MIDAS-U0 for different numbers of static and dynamic factors r and q as well as information criteria selection, MSE relative to GDP variance and ranking

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1.a. MIDAS-U0	$r=1,\ q=1$	0.71	0.86	0.89	0.90	1.05	0.98	1.05	1.09	1.12
	$r=2,\ q=1$	0.75	0.82	0.87	0.78	1.01	0.94	1.28	1.13	1.15
	$r=2,\ q=2$	0.76	0.83	0.88	0.80	1.02	0.96	1.30	1.14	1.15
	$r=3,\ q=1$	0.75	0.89	0.96	0.87	1.13	0.88	1.54	1.17	1.11
	$r=3,\ q=2$	0.72	0.87	0.89	0.82	1.02	0.87	1.46	1.11	1.00
	$r=3,\ q=3$	0.75	0.86	0.89	0.84	1.06	0.95	1.64	1.16	1.08
	$r=4,\ q=1$	0.85	0.93	1.08	1.08	1.16	1.22	1.50	1.25	1.17
	$r=4,\ q=2$	0.74	0.87	1.13	0.83	1.08	1.02	1.48	1.37	1.38
	$r=4,\ q=3$	0.82	1.03	1.20	0.92	1.17	1.31	1.59	1.34	1.37
	$r=5,\ q=1$	0.86	1.11	1.20	1.09	1.13	1.19	1.52	1.25	1.06
	$r=5,\ q=2$	0.83	1.23	1.10	1.05	1.13	1.24	1.51	1.55	1.17
	$r=5,\ q=3$	0.83	0.97	1.29	1.20	1.17	1.22	1.69	1.34	1.14
	$r=6,\ q=1$	0.92	1.15	1.22	1.12	1.10	1.03	1.59	1.02	1.20
	$r=6,\ q=2$	0.91	1.31	1.34	1.28	1.29	1.26	1.82	1.59	1.90
	$r=6,\ q=3$	0.91	1.34	1.15	1.11	1.11	1.63	1.96	1.59	1.37
	IC	0.70	0.89	0.89	0.82	1.03	0.93	1.56	1.15	1.05
1.b. Ranking	$r=1,\ q=1$	2	4	4	8	5	7	1	2	6
	$r=2,\ q=1$	6	1	1	1	1	4	2	4	9
	$r=2,\ q=2$	8	2	2	2	3	6	3	5	8
	$r=3,\ q=1$	5	7	7	7	11	2	9	8	5
	$r=3,\ q=2$	3	6	3	3	2	1	4	3	1
	$r=3,\ q=3$	7	3	6	6	6	5	13	7	4
	$r=4,\ q=1$	12	9	8	11	13	12	6	9	11
	$r=4,\ q=2$	4	5	10	5	7	8	5	13	15
	$r=4,\ q=3$	9	11	12	9	15	15	12	12	14
	$r=5,\ q=1$	13	12	13	12	12	10	8	10	3
	$r=5,\ q=2$	10	14	9	10	10	13	7	14	10
	$r=5,\ q=3$	11	10	15	15	14	11	14	11	7
	$r=6,\ q=1$	16	13	14	14	8	9	11	1	12
	$r=6,\ q=2$	15	15	16	16	16	14	15	15	16
	$r=6,\ q=3$	14	16	11	13	9	16	16	16	13
	IC	1	8	5	4	4	3	10	6	2

Note: See table 1 in the main text.

Table 8: Nowcast and forecast results for EM-PCA factors and MIDAS-U0 for different numbers of static factors r as well as information criteria selection, MSE relative to GDP variance and ranking

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1.a. MIDAS-U0	$r=1$	0.58	0.65	0.72	0.92	0.93	0.79	1.10	1.10	1.05
	$r=2$	0.66	1.07	0.85	0.98	0.96	0.73	1.26	1.00	2.30
	$r=3$	0.65	1.19	0.80	0.88	0.95	0.86	1.31	1.15	2.27
	$r=4$	1.08	1.48	0.88	1.21	1.33	0.86	2.42	1.35	2.41
	$r=5$	1.64	1.17	1.18	1.49	1.56	1.71	2.47	1.00	3.89
	$r=6$	1.23	1.18	1.74	2.15	1.41	2.16	2.25	0.95	3.76
	IC	1.63	1.43	1.10	1.51	1.66	1.41	2.48	1.10	3.56
1.b. Ranking	$r=1$	1	1	1	2	1	2	1	4	1
	$r=2$	3	2	3	3	3	1	2	3	3
	$r=3$	2	5	2	1	2	3	3	6	2
	$r=4$	4	7	4	4	4	4	5	7	4
	$r=5$	7	3	6	5	6	6	6	2	7
	$r=6$	5	4	7	7	5	7	4	1	6
	IC	6	6	5	6	7	5	7	5	5

Note: See table 1 in the main text.

Table 9: Nowcast and forecast results for KFS-PCA factors and MIDAS-U0 for different numbers of static and dynamic factors r and q as well as information criteria selection, MSE relative to GDP variance and ranking

		nowcast current quarter			forecast 1 quarter			forecast 2 quarters		
	horizon h_m	1	2	3	4	5	6	7	8	9
1.a. MIDAS-U0	$r=1,\ q=1$	0.68	0.85	0.80	0.95	1.01	0.93	1.08	1.09	1.06
	$r=2,\ q=1$	0.71	1.06	0.87	0.94	0.96	0.69	1.17	1.11	1.52
	$r=2,\ q=2$	0.66	0.99	0.83	0.94	0.97	0.69	1.19	1.09	1.52
	$r=3,\ q=1$	1.42	1.01	1.07	0.97	0.66	0.83	1.16	1.39	2.36
	$r=3,\ q=2$	0.83	1.02	1.04	0.99	0.98	0.68	1.19	1.16	1.53
	$r=3,\ q=3$	0.78	1.00	0.96	0.92	0.93	0.70	1.21	1.21	1.64
	$r=4,\ q=1$	1.74	1.31	1.12	0.78	1.09	1.02	1.54	2.07	1.99
	$r=4,\ q=2$	1.60	1.19	1.18	1.09	1.14	0.87	1.01	2.17	1.70
	$r=4,\ q=3$	1.28	1.17	1.04	1.33	1.14	0.78	1.72	1.45	1.54
	$r=5,\ q=1$	2.00	1.61	0.99	0.95	1.25	1.22	1.78	2.71	1.88
	$r=5,\ q=2$	1.79	1.32	1.32	1.21	1.20	0.88	1.17	2.39	1.76
	$r=5,\ q=3$	1.27	1.02	1.25	1.56	0.98	1.05	1.83	1.45	1.75
	$r=6,\ q=1$	1.90	1.61	1.01	1.18	1.34	1.80	1.94	2.90	2.70
	$r=6,\ q=2$	1.76	1.47	1.52	1.47	1.33	1.31	1.91	2.45	3.25
	$r=6,\ q=3$	0.96	1.41	1.55	1.72	1.10	1.42	1.99	1.38	2.20
	IC	1.51	1.12	1.01	1.27	1.04	0.80	1.61	1.31	1.69
1.b. Ranking	$r=1,\ q=1$	2	1	1	6	7	10	2	1	1
	$r=2,\ q=1$	3	7	3	4	3	2	4	3	3
	$r=2,\ q=2$	1	2	2	3	4	3	7	2	2
	$r=3,\ q=1$	9	4	10	7	1	7	3	8	14
	$r=3,\ q=2$	5	6	8	8	5	1	6	4	4
	$r=3,\ q=3$	4	3	4	2	2	4	8	5	6
	$r=4,\ q=1$	12	11	11	1	9	11	9	11	12
	$r=4,\ q=2$	11	10	12	9	11	8	1	12	8
	$r=4,\ q=3$	8	9	9	13	12	5	11	9	5
	$r=5,\ q=1$	16	16	5	5	14	13	12	15	11
	$r=5,\ q=2$	14	12	14	11	13	9	5	13	10
	$r=5,\ q=3$	7	5	13	15	6	12	13	10	9
	$r=6,\ q=1$	15	15	6	10	16	16	15	16	15
	$r=6,\ q=2$	13	14	15	14	15	14	14	14	16
	$r=6,\ q=3$	6	13	16	16	10	15	16	7	13
	IC	10	8	7	12	8	6	10	6	7

Note: See table 1 in the main text.