Factor nowcasting of German GDP with ragged-edge data: A model comparison using MIDAS projections

Massimiliano Marcellino
Università Bocconi, IGIER and CEPR

Christian Schumacher
Deutsche Bundesbank

25 October 2007
1 Introduction

Bundesbank involved in disaggregated Euro area projection exercises (BMPE) together with European Central Bank (ECB) and other National Central Banks of the Eurosystem

In BMPE, Bundesbank has to provide forecasts for German GDP and inflation

Forecasts are monitored by other Eurosystem central banks and the public

→ There is always the need to be up-to-date with forecasting techniques
Presentation today part of bigger Bundesbank project on development of a range of nowcast and forecast models

Techniques are still under evaluation, not implemented yet in regular forecast exercises
Outline of presentation:

1. What is nowcasting?

2. Difficulties of now- and forecasting GDP

3. Discussion of three factor models that can tackle unbalanced data

4. MIDAS projection with factors as regressors: FACTOR-MIDAS

5. Empirical now- and forecast comparison

6. Conclusions
2 What is nowcasting?

Decision makers regularly request information on the current state of the economy.

GDP is an important business cycle indicator, but sampled at quarterly frequency only and published with considerable delay.

Economist’s task: **Estimate current quarter GDP** using all information which is currently available.

Example: in April, German GDP is available only for the fourth quarter of the previous year. To obtain a 2\textsuperscript{nd} quarter GDP nowcast, we have to make a projection with forecast horizon of two quarters from the end of the GDP sample.
3 Difficulties of now- and forecasting GDP

There are many difficulties in real-world applications, we discuss two:

1. GDP is quarterly data, many important indicators are sampled at monthly or higher frequency - mixed-frequency problem

2. Indicators for nowcasting are available with different publication lags → leads to the so-called ragged edge of multivariate datasets, Wallis (1986)

Question here: How to nowcast quarterly GDP with factors estimated from a large set of monthly indicators?
4 Factor now- and forecasting: Two-step procedure

Boivin and Ng (2005) IJCB: In the single-frequency case, factor forecasting is often a two-step procedure

1. Estimation Step: Estimate factors

2. Forecast step: Factor-augmented (V)AR model

We follow the same strategy in the mixed-frequency case with ragged-edge data
5 Factor estimation with ragged-edge data

Monthly observations have a factor structure

\[ X_{tm} = \Lambda F_{tm} + \xi_{tm} \]  \hspace{1cm} (1)

with factors \( F_{tm} = (f_{1,tm}', \ldots, f_{r,tm}') \), loadings \( \Lambda \), idiosyncratic components \( \xi_{tm} \)

If data \( X = (X_1, \ldots, X_{T_m})' \) is balanced, there are different ways to estimate \( F \):

1. PCA as in Stock and Watson (2002) JBES

2. Dynamic PCA according to Forni et al. (2005) JASA

We consider three factors models for ragged-edge data:

1. Vertical realignment of the data and dynamic PCA, Altissimo et al. (2006) CEPR WP (New Eurocoin)


3. Kalman smoother estimation using a large state-space factor model with PCA providing initial values, see Doz, Giannone and Reichlin (2006) ECB WP
5.1 Vertical realignment of the data and dynamic PCA

Altissimo et al. (2006) CEPR WP

Variable $i$ is released with $k_i$ months of publication lag → in period $T_m$, the final observation available is in period $T_m - k_i$

Balancing by ‘vertical’ realignment

$$
\tilde{x}_{i,T_m} = x_{i,T_m-k_i}
$$

(2)

Applying this procedure for each series and harmonising at the beginning of the sample, yields a balanced data set $\tilde{X}_{t_m}$

Factor estimation from $\tilde{X}_{t_m}$ by Dynamic PCA, Forni et al. (2005) JASA
5.2 EM algorithm and static PCA


We want a full data column vector $\mathbf{X}_i = (x_{i,1}, \ldots, x_{i,T_m})'$, due to the ragged-edge problem we observe only $\mathbf{X}_i^{\text{obs}}$

relationship between observed and unobserved

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

(3)

where $\mathbf{A}_i$ is a matrix that can tackle missing values or mixed frequencies
EM algorithm:

1. Initial (naive) guess of observations $\hat{X}_i^{(0)} \forall i$ yields a balanced dataset $\hat{X}^{(0)}$, static PCA provides initial monthly factors $\hat{F}^{(0)}$ and loadings $\hat{\Lambda}^{(0)}$

2. **E-step:** Expectation of $X_i$ conditional on observations $X_i^{\text{obs}}$, factors $\hat{F}^{(j-1)}$ and loadings $\hat{\Lambda}_i^{(j-1)}$ from the previous iteration

   \[
   \hat{X}_i^{(j)} = \hat{F}^{(j-1)}\hat{\Lambda}_i^{(j-1)} + A_i'(A_i'A_i)^{-1} \left( X_i^{\text{obs}} - A_i\hat{F}^{(j-1)}\hat{\Lambda}_i^{(j-1)} \right) \tag{4}
   \]

3. **M-step:** Repeat the E-step for all $i$ providing again a balanced dataset. Reestimate $\hat{F}^{(j)}$ and $\hat{\Lambda}^{(j)}$ by PCA, and go to step 2 until convergence
5.3 Kalman smoother estimates in a large state-space model

Doz et al. (2006) ECB WP

Model:

\[ X_{tm} = \Lambda F_{tm} + \xi_{tm} \]  
(5)

\[ \psi(L_m)F_{tm} = B\eta_{tm} \]  
(6)

VAR for factors with \( \psi(L_m) = \sum_{i=1}^{p} \psi_i L_m^i \) and \( L_m x_{tm} = x_{tm-1} \), \( q \)-dimensional vector \( \eta_{tm} \) contains dynamic shocks that drive factors, identification matrix \( B \) is \( (r \times q) \)-dimensional

Model has state-space representation with factors as states

Estimation ‘trick’: Coefficients estimated outside state-space model, no iterative ML
QML to estimate the factors:

1. Estimate $\hat{F}_{tm}$ using PCA as an initial estimate

2. Estimate $\hat{\Lambda}$ by regressing $X_{tm}$ on the estimated factors $\hat{F}_{tm}$. The covariance of the idiosyncratic components $\hat{\xi}_{tm} = X_{tm} - \hat{\Lambda}\hat{F}_{tm}$, denoted as $\hat{\Sigma}_\xi$, is also estimated

3. Estimate factor VAR($p$) on the factors $\hat{F}_{tm}$ yielding $\hat{\Psi}(L_m)$ and the residual covariance of $\hat{\zeta}_{tm} = \hat{\Psi}(L_m)\hat{F}_{tm}$, denoted as $\hat{\Sigma}_\zeta$

4. To obtain an estimate for $\mathbf{B}$, apply an eigenvalue decomposition of $\hat{\Sigma}_\zeta$

5. State-space is fully specified numerically $\rightarrow$ The Kalman smoother provides estimates of factors
5.4 Discussion

VA-DPCA: easy to use, but

- availability of data determines dynamic cross-correlations between variables!

- statistical release dates for data are not the same over time, for example due to major revisions → dynamic correlations within the data change and factors can change over time

EM-PCA

- interpolation of missing values in line with factor model
- no dynamics

- sometimes convergence problems ('partial factors', large $r$)

KFS-PCA

- optimality properties of Kalman filter/smooth

- explicit dynamics of the factors

- assumptions on idiosyncratic components often not fulfilled

- more (dynamic) structure, more auxiliary parameters to fix and perhaps subject to misspecification, see Boivin and Ng (2005) IJCB
6 Factor forecasting: MIDAS-basic


Our contribution: Clements and Galvão (2007) WP use single macro variables, we use factors → FACTOR-MIDAS

Three ways:

1. MIDAS-basic

2. MIDAS-smooth

3. Unrestricted MIDAS
6.1 Factor forecasting: MIDAS-basic


MIDAS Model with one factor $\hat{f}_{tm}$ for forecast horizon $hq$ quarters ($hq = hm/3$):

$$
  y_{tq+hq} = \beta_0 + \beta_1 b(L_m, \theta)\hat{f}_{tm}^{(3)} + \varepsilon_{tm+h_q}
$$

(7)

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

(8)

Quarterly GDP $y_{tq+hq}$ is directly related to the monthly factor $\hat{f}_{tm}^{(3)}$ and lags, where $\hat{f}_{tm}^{(3)}$ is a skip-sampled version of $\hat{f}_{tm}$. 
Estimation by nonlinear least squares (NLS)

Device for **direct** forecasting depending on $h_q$, see Marcellino, Stock, Watson (2006)

$$y_{T_q+h_q|T_q} = \hat{\beta}_0 + \hat{\beta}_1 b(L_m, \hat{\theta})\hat{f}_{T_m}$$  \hspace{1cm} (9)
6.2 The projection in New Eurocoin: MIDAS-smooth

Altissimo et al. (2006) CEPR WP

Projection:

\[ y_{T_q+h_q|T_q} = \hat{\mu} + \left[ \tilde{\Sigma}_y F(h_m) \times \tilde{\Sigma}_F^{-1} \right] \times \tilde{F}_{T_m} \] (10)

\( \hat{\mu} \) is GDP sample mean, \( \tilde{\Sigma}_F \) is sample covariance of factors

\( \tilde{\Sigma}_y F(k) \) is cross-covariance between smoothed GDP and factors with \( k \) lags

\[ \tilde{\Sigma}_y F(k) = \frac{1}{2H + 1} \sum_{j=-H}^{H} \alpha(\omega_j) \hat{S}_y F(\omega_j) e^{i\omega_j k} \] (11)

with \( \alpha(\omega_j) = 1 \forall \left| \omega_j \right| < \pi/6 \) and zero otherwise, and cross-spectral matrix \( \hat{S}_y F(\omega_j) \)
If we disregard smoothing and use

$$\hat{\Sigma}_{yF}(k) = \frac{1}{T^* - 1} \sum_{t_m = M + 1}^{T_m} y_{t_m} \hat{F}_{t_m-k}^{(3)'}$$

(12)

where $T^* = \text{floor}[(T_m - (M + 1))/3]$ in

$$y_{T_q+h_q|T_q} = \hat{\mu} + \left[ \hat{\Sigma}_{yF}(h_m) \times \hat{\Sigma}_F^{-1} \right] \times \hat{F}_{T_m}$$

(13)

→ Both in basic MIDAS and New Eurocoin: $\hat{\Sigma}_{yF}(k)$ can be estimated consistently although $y_{t_q}$ and $\hat{F}_{t_m}$ have different sampling frequencies

→ Projection in New Eurocoin and basic MIDAS projection follow the same idea!
6.3 Unrestricted MIDAS

Unrestricted lag order model

\[ y_{t_q+h_q} = \beta_0 + D(L_m)\hat{F}_{t_m}^{(3)} + \varepsilon_{t_m+h_m}, \]  

(14)

where \( D(L_m) = \sum_{k=0}^{K} D_k L_m^k \) is an unrestricted lag polynomial of order \( K \).

Estimation of \( D(L_m) \) and \( \beta_0 \) by OLS

We consider fixed lag orders with \( k = 0 \) and \( k = 1 \)

Note that for \( k = 0 \), we consider only \( t_m \)-dated factors for forecasting \( \rightarrow \) case \( k = 0 \) is like MIDAS-smooth without smoothing
7 Empirical now- and forecast comparison

7.1 Data

German quarterly GDP from 1992Q1 until 2006Q3

111 monthly indicators

Data: No real-time, we generate ragged-edge vintages from final data

Solution: Missing values at the end of the full, final sample are used to identify missing values in pseudo real-time subsamples, see Banbura and Rünstler (2007) ECB WP
7.2 Forecast design

Recursive design with increasing sample size

Evaluation sample from 1998Q4 until 2006Q3

Each month, we compute new now- and forecasts with monthly horizon $h_m = 1, 2, \ldots, 9$

Models are estimated using fixed specification in terms of numbers of factors $r$ and $q$, and using information criteria, see Bai and Ng (2002, 2007)
7.3 Empirical results

Results based on fixed specifications for $r = 1, 2$

<table>
<thead>
<tr>
<th></th>
<th>horizon $h_m$</th>
<th>nowcast</th>
<th>forecast</th>
<th>forecast</th>
<th>forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>current quarter</td>
<td>1 quarter</td>
<td>2 quarters</td>
<td>3 quarters</td>
</tr>
<tr>
<td>1.a. MIDAS-U0</td>
<td></td>
<td>VA-DPCA</td>
<td>0.71</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EM-PCA</td>
<td>0.58</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KFS-PCA</td>
<td>0.68</td>
<td>0.85</td>
<td>0.80</td>
</tr>
<tr>
<td>1.b. Ranking</td>
<td></td>
<td>VA-DPCA</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EM-PCA</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KFS-PCA</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2. Benchmarks</td>
<td></td>
<td>AR</td>
<td>1.02</td>
<td>1.17</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in-sample mean</td>
<td>1.03</td>
<td>1.04</td>
<td>1.04</td>
</tr>
</tbody>
</table>
relative MSE to variance of GDP, number of factors $r = 2, q = 1$

<table>
<thead>
<tr>
<th></th>
<th>horizon $h_m$</th>
<th>nowcast</th>
<th>forecast</th>
<th>forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>current quarter</td>
<td>1 quarter</td>
<td>2 quarters</td>
</tr>
<tr>
<td>4.a. MIDAS-U0</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>VA-DPCA</td>
<td>0.75</td>
<td>0.82</td>
<td>0.87</td>
<td>0.78</td>
</tr>
<tr>
<td>EM-PCA</td>
<td>0.66</td>
<td>1.07</td>
<td>0.85</td>
<td>0.98</td>
</tr>
<tr>
<td>KFS-PCA</td>
<td>0.71</td>
<td>1.06</td>
<td>0.87</td>
<td>0.94</td>
</tr>
<tr>
<td>4.b. Ranking</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>VA-DPCA</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>EM-PCA</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>KFS-PCA</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>6. Benchmarks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>1.02</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
</tr>
<tr>
<td>in-sample mean</td>
<td>1.03</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
</tr>
</tbody>
</table>
### relative MSE to variance of GDP, number of factors $r = 1$

<table>
<thead>
<tr>
<th></th>
<th>horizon $h_m$</th>
<th>nowcast current quarter</th>
<th>forecast 1 quarter</th>
<th>forecast 2 quarters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>1.a. VA-DPCA</strong></td>
<td>MIDAS-basic</td>
<td>0.71</td>
<td>1.01</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>MIDAS-smooth</td>
<td>0.69</td>
<td>0.92</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>MIDAS-U0</td>
<td>0.71</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td><strong>1.b. Ranking</strong></td>
<td>MIDAS-basic</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>MIDAS-smooth</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MIDAS-U0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>2.a. EM-PCA</strong></td>
<td>MIDAS-basic</td>
<td>0.62</td>
<td>0.69</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>MIDAS-smooth</td>
<td>0.70</td>
<td>0.73</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>MIDAS-U0</td>
<td>0.58</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>2.b. Ranking</strong></td>
<td>MIDAS-basic</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>MIDAS-smooth</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>MIDAS-U0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
7.4 More results

Nowcasts based on information-criteria model selection for \( r \) and \( q \) have no information content.

We also checked PCA vs DPCA together with vertical realignment of the data \( \rightarrow \) no big differences.

We checked the information content of ragged-edge data vs. balanced data \( \rightarrow \) ragged-edge contains in general useful information for nowcasting.

We compared an integrated state-space model which also interpolates GDP from Bambura and Rünstler (2007) with the two-step factor nowcast here \( \rightarrow \) no big differences.
8 Conclusions

Factor models considered here can address nowcasting questions with ragged-edge and mixed-frequency data.

Models with only one or two factors \((r = 1, 2)\) perform best.

Differences between factor estimation methods (vertical realignment, EM, state space) are minor.

Simplest MIDAS projections with few lags do better than exponential lag versions.

Projections are informative for the nowcast and forecast one quarter ahead → factor models can be regarded as **short-term forecast models** only.