FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure

Frale C., Monteforte L.

Computational and Financial Econometrics
Limassol, October 2009

Introduction

- After the recent financial and economic crisis there is an increasing new demand for macroeconometric models able to predict the state of the economy and to capture early signals of turning points.
- Classical models for short term forecast, such as bridge models and standard factors models, have shown some limitations, especially as regard as the time aggregation and the ragged-edge data problem.
- New approaches, such as mixed frequency factor models and MIDAS regressions are suitable for solving this two issues

Motivation and main results

- We combine this two approaches, mixed frequency factors and MIDAS, in order to exploit in a parsimonious way a larger number of lags in a multivariate framework. This is particulary useful in forecasting as it allows to explicitly take into account the cross correlation between indicators and the target variable with different frequencies.
- Moreover the MIDAS polynomial produces smooth factors and less volatile forecasts.
- We compare the forecasting power of our model (FaMIDAS) with a number of competing models. We find that it tends to prevail at larger horizons in real time experiment.

Outline

- ✓ Introduction and motivation
- ✓ The Model
  - The factor model with mixed frequency
  - The MIDAS for the lags combination
  - The combination: FaMIDAS
- ✓ Empirical Application
- ✓ Forecasting performance
- ✓ Conclusion and future agenda

1. The factor model with mixed frequency

We refer to the Monthly Indicator of the economic activity in the Euro Area, developed by Eurostat and documented in Frale et al.(2008):

$$
\begin{array}{lcll}
\mathbf{y}_t & = & \vartheta_{\mathbf{0}}\mu_t + \vartheta_{\mathbf{1}}\mu_{t-1} + \widetilde{\vartheta_{\mathbf{0}}}\widetilde{\mu}_t + \widetilde{\vartheta_{\mathbf{1}}}\widetilde{\mu}_{t-1} + \boldsymbol{\gamma}_t + \mathbf{X}_t\beta, & t = 1, \dots, n, \\
\phi(L)\Delta\mu_t & = & (1-\theta L)^p\eta_t, & \eta_t \sim \mathrm{NID}(0, \sigma^2_\eta), \\
\widetilde{\phi}(L)\Delta\widetilde{\mu}_t & = & \widetilde{\eta}_t, & \widetilde{\eta}_t \sim \mathrm{NID}(0, \sigma^2_{\widetilde{\eta}}), \\
\mathbf{D}(L)\Delta\boldsymbol{\gamma}_t & = & \boldsymbol{\delta} + \boldsymbol{\xi}_t, & \boldsymbol{\xi}_t \sim \mathrm{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\xi),
\end{array}
$$

$(1-\theta L)^p\eta_t$ is the pre-specified MA(p) term which squeezes the spectrum in the interval $(1-\theta)/(1+\theta)$ and therefore accounts for low frequency cycles (Morton & Tunnicliffe-Wilson, G. (2000)).

$\phi(L)$ and $\widetilde{\phi}(L)$ are autoregressive polynomials of order p and $\widetilde{p}$ with stationary roots
The matrix polynomial $\mathbf{D}(L)$ is diagonal and $\Sigma_\xi = \mathrm{diag}(\sigma^2_1, \dots, \sigma^2_N)$.

The disturbances η_t ,$\widetilde{\eta}_t$ and ξ_t are mutually uncorrelated at all leads and lags.

1. The factor model with mixed frequency

We refer to the Monthly Indicator of the economic activity in the Euro Area, developed by Eurostat and documented in Frale et al.(2008):

$$
\begin{array}{rcll}
\mathbf{y}_t & = & \vartheta_{\mathbf{0}}\mu_t + \vartheta_{\mathbf{1}}\mu_{t-1} + \widetilde{\vartheta_{\mathbf{0}}}\widetilde{\mu}_t + \widetilde{\vartheta_{\mathbf{1}}}\widetilde{\mu}_{t-1} + \boldsymbol{\gamma}_t + \mathbf{X}_t\beta, & t = 1, ..., n, \\
\phi(L)\Delta\mu_t & = & (1-\theta L)^p \eta_t, & \eta_t \sim \mathrm{NID}(0, \sigma^2_\eta), \\
\widetilde{\phi}(L)\Delta\widetilde{\mu}_t & = & \widetilde{\eta}_t, & \widetilde{\eta}_t \sim \mathrm{NID}(0, \sigma^2_{\widetilde{\eta}}), \\
\mathbf{D}(L)\Delta\boldsymbol{\gamma}_t & = & \boldsymbol{\delta} + \boldsymbol{\xi}_t, & \boldsymbol{\xi}_t \sim \mathrm{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\xi),
\end{array}
$$

$(1-\theta L)^p \eta_t$ is the pre-specified MA(p) term which squeezes the spectrum in the interval $(1-\theta)/(1+\theta)$ and therefore accounts for low frequency cycles (Morton & Tunnicliffe-Wilson, G. (2000)).

$\phi(L)$ and $\widetilde{\phi}(L)$ are autoregressive polynomials of order p and $\widetilde{p}$ with stationary roots
The matrix polynomial $\mathbf{D}(L)$ is diagonal and $\Sigma_\xi = \mathrm{diag}(\sigma^2_1, \ldots, \sigma^2_N)$.

The disturbances η_t ,$\widetilde{\eta}_t$ and ξ_t are mutually uncorrelated at all leads and lags.

Estimation and time constraint procedure

- The model involves mixed frequency data, e.g. monthly indicators and quarterly GDP. Following Harvey (1989) and Proietti(2006), the state vector in the SSF is suitably augmented by using an appropriately defined cumulator variable in order to traslate the time constraint into a problem of missing observations.
- The model is cast in State Space Form and, under Gaussian distribution of the errors, the unknown parameters can be estimated by maximum likelihood, using the prediction error decomposition, performed by the Kalman filter.
- Filter and Smoother are based on the Univariate statistical treatment of multivariate models by Koopman and Durbin (2000): very flexible and convenient device for handling missing values in multivariate models and reduce the time of convergence.

2. The MIDAS for the lags combination

- The anticipating power of an economic series for any target variable is purely an empirical aspect, even more cumbersome with mixed frequency data. An efficient and suitable solution are MIDAS models that summarize and combine the information content of the indicators and their lags, with weights jointly estimated.

- A MIDAS regression takes the form:

$$Y_t = \beta_0 + B(\theta, L^{1/m})X_t^m + \varepsilon_t$$

where $B(\theta, L_{1/m}) = \sum_{k=0}^{k} b(\theta, k)L^{k/m}$ is a polynomial of lag k and $L^{1/m}$ is an operator such that $L^{k/m}X_t^m = X_{t-k/m}^m$. In other words the regression equation is a projection of Y_t into a higher frequency series X_t^m up to k lags back.

2. The MIDAS for the lags combination (cont.)

The most common weights structure are:

- a parametrization that refers to Almon lags:

$$b(k;\theta) = \frac{exp(\theta_1 k + \dots \theta_q k^q)}{\sum_{j=1}^{k} (\theta_1 k + \dots \theta_q k^q)}.$$

  The simplicity of the Almon weights might be preferable in the case of small number of time lags involved

- weights drown by a Beta distribution, such as:

$$b(k;\theta_1,\theta_2) = \frac{f(k;\theta_1,\theta_2)}{\sum_{j=1}^{k} f(k;\theta_1,\theta_2)}$$

  where $f(x,a,b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$, $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma(a) = \int_0^\infty e^(-x)x^{a-1}dx$.

The FaMIDAS

The FaMIDAS results by the following equations:

$$
\begin{aligned}
\begin{bmatrix} b(L_k,\theta)\mathbf{x_t} \\ \mathbf{y}_{2,t} \end{bmatrix} &= \vartheta_0 f_t + \gamma_t + \mathbf{S}_t\beta, & t = 1,\dots,n, \\
\phi(L)\Delta f_t &= \eta_t, & \eta_t \sim \mathrm{NID}(0,\sigma_\eta^2), \\
\mathbf{D}(L)\Delta\gamma_t &= \delta + \eta_t^*, & \eta_t^* \sim \mathrm{NID}(\mathbf{0},\Sigma_{\eta^*}),
\end{aligned}
$$

In our application $b(L_k,\theta)$ is the exponential Almon lag polynomial: $\sum_{k=0}^{K} w(k,\theta)L^k$ with

$$
w(k,\theta) = \frac{exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=0}^{K} exp(\theta_1 k + \theta_2 k^2)}.
$$

The FaMIDAS

The FaMIDAS results by the following equations:

MIDAS

$$
\begin{aligned}
\begin{bmatrix} b(L_k,\theta)\mathbf{x}_t \\ \mathbf{y}_{2,t} \end{bmatrix} &= \vartheta_0 f_t + \gamma_t + \mathbf{S}_t\beta, \quad & t = 1,\dots,n, \\
\phi(L)\Delta f_t &= \eta_t, & \eta_t \sim \text{NID}(0,\sigma_\eta^2), \\
\mathbf{D}(L)\Delta\gamma_t &= \delta + \eta_t^*, & \eta_t^* \sim \text{NID}(\mathbf{0},\Sigma_{\eta^*}),
\end{aligned}
$$

In our application $b(L_k,\theta)$ is the exponential Almon lag polynomial: $\sum_{k=0}^{K} w(k,\theta)L^k$ with

$$
w(k,\theta) = \frac{exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=0}^{K} exp(\theta_1 k + \theta_2 k^2)}.
$$

Figure: Monthly Indicators and Quarterly GDP- Italy

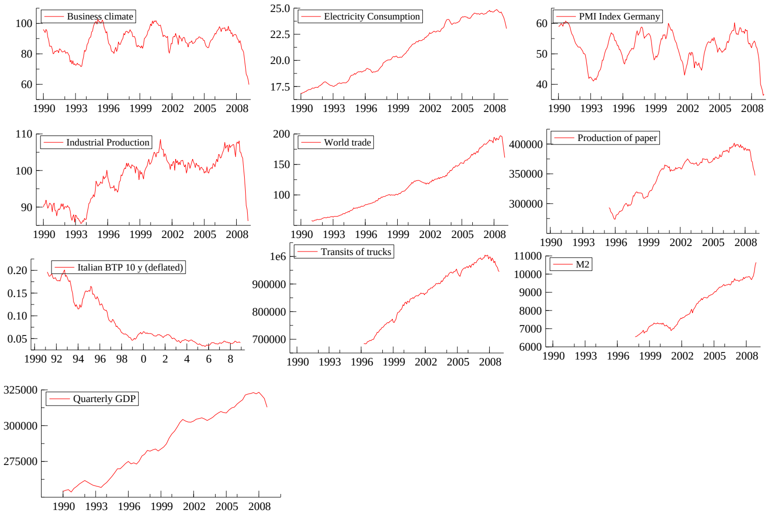


Table: Maximum likelihood estimated factor loadings ϑ-(1990M1-2009M4)

	MIXFAC	**MIX2FAC**		**FaMIDAS**
		Factor 1	Factor 2	weights
ISAE Business Climate	0.44 **	-0.61 **	-0.02	0.09 **
Electricity Consumption	0.01	-0.03 **	0.01	0.05 **
PMI Germany	0.35 *	-0.46*	-0.12	0.06 **
Industrial production	0.44 **	-0.53 **	0.10	0.06 **
GDP	0.16 **	-0.17 **	0.01	0.02 **
PMI Germany(-1)	-0.22			
IP(-1)	0.67 **			
Industrial production of paper		-0.14 **	0.03	
World trade (CPB)		-0.74 **	0.17	
Italian Treasury bonds yield (10y)		-0.03	-0.37**	
Money supply		0.24 **	-0.02	
Motorway flows (trucks)		-0.17 *	0.01	

** means significant at 5%, * at 10%.

Business Climate is provided by ISAE; Electricity is the monthly consumption of electricity provided by TERNA; PMI Germany is the Purchase Manager Index for Germany in manufacturing and services; IP paper is the Industrial production of paper and cardboard; World trade is the indicator of trade produced by the CPB- Netherlands Bureau for Economic Policy Analysis; Money supply includes currency and deposits; Motorway flow refers to trucks and is provided by Autostrade

Table: Maximum likelihood estimated factor loadings ϑ-(1990M1-2009M4)

	MIXFAC	MIX2FAC		FaMIDAS
		Factor 1	Factor 2	weights
ISAE Business Climate	0.44 **	-0.61 **	-0.02	0.09 **
Electricity Consumption	0.01	-0.03 **	0.01	0.05 **
PMI Germany	0.35 *	-0.46*	-0.12	0.06 **
Industrial production	0.44 **	-0.53 **	0.10	0.06 **
GDP	0.16 **	-0.17 **	0.01	0.02 **
PMI Germany(-1)	-0.22			
IP(-1)	0.67 **			
Industrial production of paper		-0.14 **	0.03	
World trade (CPB)		-0.74 **	0.17	
Italian Treasury bonds yield (10y)		-0.03	-0.37**	
Money supply		0.24 **	-0.02	
Motorway flows (trucks)		-0.17 *	0.01	

** means significant at 5%, * at 10%.

Business Climate is provided by ISAE; Electricity is the monthly consumption of electricity provided by TERNA; PMI Germany is the Purchase Manager Index for Germany in manufacturing and services; IP paper is the Industrial production of paper and cardboard; World trade is the indicator of trade produced by the CPB- Netherlands Bureau for Economic Policy Analysis; Money supply includes currency and deposits; Motorway flow refers to trucks and is provided by Autostrade

Table: Maximum likelihood estimated factor loadings ϑ-(1990M1-2009M4)

	MIXFAC	MIX2FAC Factor 1	MIX2FAC Factor 2	FaMIDAS
				weights
ISAE Business Climate	0.44 **	-0.61 **	-0.02	0.09 **
Electricity Consumption	0.01	-0.03 **	0.01	0.05 **
PMI Germany	0.35 *	-0.46*	-0.12	0.06 **
Industrial production	0.44 **	-0.53 **	0.10	0.06 **
GDP	0.16 **	-0.17 **	0.01	0.02 **
PMI Germany(-1)	-0.22			
IP(-1)	0.67 **			
Industrial production of paper		-0.14 **	0.03	
World trade (CPB)		-0.74 **	0.17	
Italian Treasury bonds yield (10y)		-0.03	-0.37**	
Money supply		0.24 **	-0.02	
Motorway flows (trucks)		-0.17 *	0.01	

** means significant at 5%, * at 10%.

Business Climate is provided by ISAE; Electricity is the monthly consumption of electricity provided by TERNA; PMI Germany is the Purchase Manager Index for Germany in manufacturing and services; IP paper is the Industrial production of paper and cardboard; World trade is the indicator of trade produced by the CPB- Netherlands Bureau for Economic Policy Analysis; Money supply includes currency and deposits; Motorway flow refers to trucks and is provided by Autostrade

Figure: Estimated Monthly GDP (growth rate) and common factors for the three models.

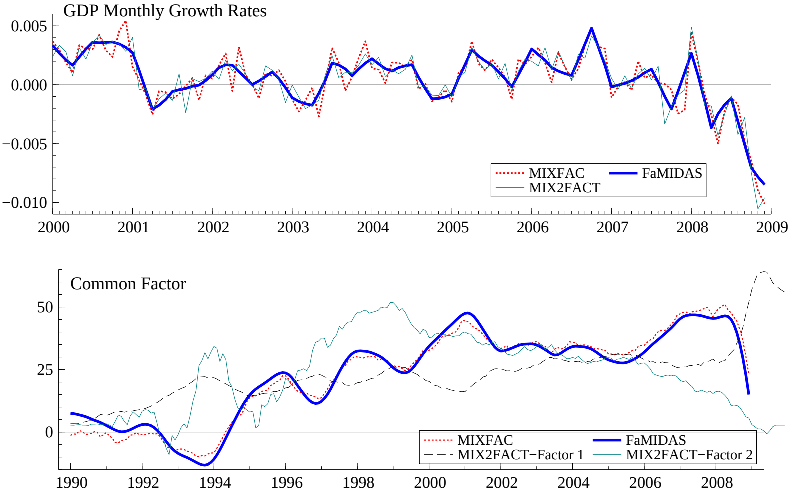
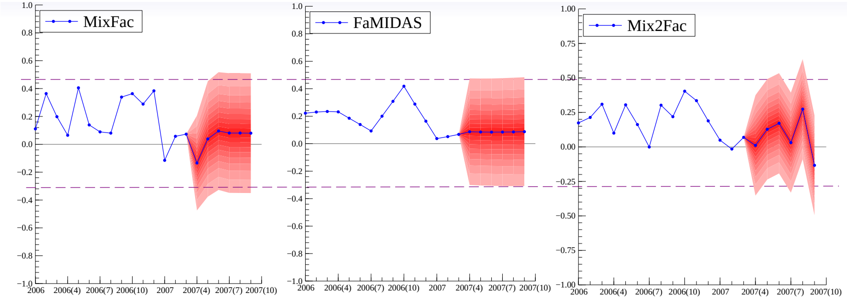


Figure: Forecasts and fan chart for the three model



Note: The filled area is the simulated 95% confidence band.

Table: Rolling RMSFE by month of the quarter, horizon of prevision and window length.

	5 years (2003-2007)			4 years (2004-2007)			3 years (2005-2007)		
VAR	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 3		0.43	0.43		0.41	0.43		0.40	0.36
ADL	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.31	0.42		0.30	0.42		0.30	0.43	
Month 2		0.40	0.45		0.40	0.46		0.41	0.50
Month 3		0.34	0.45		0.33	0.46		0.33	0.50
MIXFAC	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	$\underline{0.24}$	$\underline{0.34}$		$\underline{0.22}$	$\underline{0.32}$		0.23	$\underline{0.35}$	
Month 2		0.31	0.37		0.30	0.37		0.31	0.39
Month 3		$\underline{0.27}$	0.34		$\underline{0.25}$	$\underline{0.33}$		0.24	$\underline{0.34}$
MIX2FAC	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	$\underline{0.24}$	0.36		$\underline{0.22}$	$\underline{0.32}$		$\underline{0.22}$	0.36	
Month 2		$\underline{0.31}$	0.35		$\underline{0.29}$	$\underline{0.33}$		$\underline{0.30}$	$\underline{0.34}$
Month 3		0.34	0.36		0.30	0.34		0.29	0.35
FaMIDAS	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.32	0.36		0.32	0.35		0.36	0.37	
Month 2		0.34	$\underline{0.31}$		0.35	$\underline{0.33}$		0.37	$\underline{0.34}$
Month 3		0.34	$\underline{0.33}$		0.34	0.35		0.36	0.35
Pooling	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
	0.21	0.33		0.20	0.30		0.21	0.33	
		0.30	0.33		0.29	0.32		0.30	0.33
		0.28	0.32		0.26	0.31		0.24	0.32

Note: Best values are underlined. The VAR is estimated on a balanced quarterly sample.

Table: Rolling RMSFE by month of the quarter, horizon of prevision and window length.

	5 years (2003-2007)			4 years (2004-2007)			3 years (2005-2007)		
VAR	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 3		0.43	0.43		0.41	0.43		0.40	0.36
ADL	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.31	0.42		0.30	0.42		0.30	0.43	
Month 2		0.40	0.45		0.40	0.46		0.41	0.50
Month 3		0.34	0.45		0.33	0.46		0.33	0.50
MIXFAC	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.24	0.34		0.22	0.32		0.23	0.35	
Month 2		0.31	0.37		0.30	0.37		0.31	0.39
Month 3		0.27	0.34		0.25	0.33		0.24	0.34
MIX2FAC	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.24	0.36		0.22	0.32		0.22	0.36	
Month 2		0.31	0.35		0.29	0.33		0.30	0.34
Month 3		0.34	0.36		0.30	0.34		0.29	0.35
FaMIDAS	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.32	0.36		0.32	0.35		0.36	0.37	
Month 2		0.34	0.31		0.35	0.33		0.37	0.34
Month 3		0.34	0.33		0.34	0.35		0.36	0.35
Pooling	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
	0.21	0.33		0.20	0.30		0.21	0.33	
		0.30	0.33		0.29	0.32		0.30	0.33
		0.28	0.32		0.26	0.31		0.24	0.32

Note: Best values are underlined. The VAR is estimated on a balanced quarterly sample.

Table: Rolling RMSFE by month of the quarter, horizon of prevision and window length.

	5 years (2003-2007)			4 years (2004-2007)			3 years (2005-2007)		
VAR	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 3		0.43	0.43		0.41	0.43		0.40	0.36
ADL	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.31	0.42		0.30	0.42		0.30	0.43	
Month 2		0.40	0.45		0.40	0.46		0.41	0.50
Month 3		0.34	0.45		0.33	0.46		0.33	0.50
MIXFAC	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.24	0.34		0.22	0.32		0.23	0.35	
Month 2		0.31	0.37		0.30	0.37		0.31	0.39
Month 3		0.27	0.34		0.25	0.33		0.24	0.34
MIX2FAC	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.24	0.36		0.22	0.32		0.22	0.36	
Month 2		0.31	0.35		0.29	0.33		0.30	0.34
Month 3		0.34	0.36		0.30	0.34		0.29	0.35
FaMIDAS	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
Month 1	0.32	0.36		0.32	0.35		0.36	0.37	
Month 2		0.34	0.31		0.35	0.33		0.37	0.34
Month 3		0.34	0.33		0.34	0.35		0.36	0.35
Pooling	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}	Q_{t-1}	Q_t	Q_{t+1}
	0.21	0.33		0.20	0.30		0.21	0.33	
		0.30	0.33		0.29	0.32		0.30	0.33
		0.28	0.32		0.26	0.31		0.24	0.32

Note: Best values are underlined. The VAR is estimated on a balanced quarterly sample.

Forecasting ability

- ✓ MIXFAC outperforms the other two models for short term forecasts with small information sample.
- ✓ MIX2FAC seems to perform best for nowcasting with medium information set (second month of the quarter)
- ✓ FaMIDAS makes the lowest forecasting error for the prediction 1-quarter-ahead, given support to the intuition that it exploits efficiently the correlation of the lag structure of the indicators with the target variable.
- ✓ Pooling forecasts dominate the single models, with only few exceptions for which anyway the RMSFE is very close to the smallest.

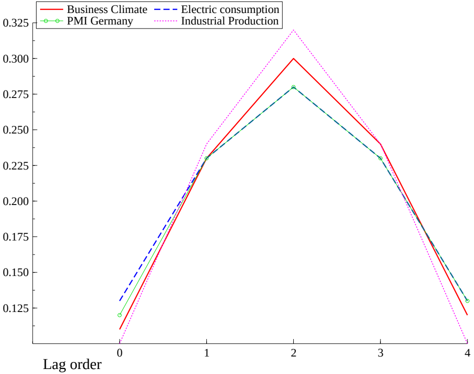
Summary and conclusion

- With the aim of improving forecasting performance, two promising direction of research are combined: The dynamic mix frequency factor models and the MIDAS structure. The latest is used in order to efficiently exploit the content of timely and high frequency macroeconomic series.
- The MIDAS structure enables to exploit in a parsimonious way a larger number of lags of the high frequency indicators, allowing to explicitly take into account the cross correlation between indicators and the target variable.
- In the empirical application the FaMIDAS produces smoother estimates for the disaggregate target variable and better forecast in a longer horizon.
- Results for this application confirm some evidence in the literature that pooling forecasts are more stable than previsions from single models.

Further research

- ✓ Extended the set of indicators (more financial variables, fiscal varaibles)
- ✓ Improving the forecast accuracy diagnostics (recursive experiment, formal tests, other benchmark, such as Marcellino & Schumacher, Ghysel et alt.)
- ✓ For the pooling consider other weight systems.

Figure: FaMIDAS estimated weights



back