FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure

Frale C., Monteforte L.

MIDAS Workshop, Goethe University Frankfurt
Frankfurt, March 4th 2010

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, especially with the aim of defining an effective economic policy.
- Classical models for short term forecast applied by Institutions, 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.

Related literature

- The mixed frequency literature with state space factor models, estimated via the Kalman filter. Most of the applications exploit monthly series, to predict the quarterly GDP. Mariano and Murasawa (2003), Mittnik and Zadrozny (2004), Aruoba et al. (2009), Camacho and Perez Quiros (2008).
- The statistical literature that uses these models as a multivariate tool for time series disaggregation, as done in Frale et al. (2008), Harvey and Chung (2000), Moauro and Savio(2005).
- The recent research Mixed Data Sampling Regression Models (MIDAS), proposed by Ghysels et al. (2002, 2006). Early applications were on financial data, more recently few applications to macroeconomic variables. Clements and Galvao (2007) and Andreou, Ghysels and Kourtellos (2010): forecasting monthly US quarterly macro variables; Ghysels and Wright (2008): tracking daily survey expectations on US macro variables; Marcellino and Schumacher (2008): monthly estimate of the German GDP.

Outline

- ✓ Introduction and motivation
- ✓ Related literature
- ✓ 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. MIXFACT-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):

$$\begin{array}{rcll} \begin{bmatrix} \mathbf{x_t} \\ y_t \end{bmatrix} & = & \vartheta_{\mathbf{0}} f_t + \vartheta_{\mathbf{1}} f_{t-1} + \boldsymbol{\gamma}_t + \mathbf{S}_t \beta, & t = 1, ..., n \\ \phi(L)\Delta f_t & = & \eta_t & \eta_t \sim \mathrm{NID}(0, \sigma^2_\eta) \\ \mathbf{D}(L)\Delta\boldsymbol{\gamma}_t & = & \boldsymbol{\delta} + \boldsymbol{\xi}_t, & \boldsymbol{\xi}_t \sim \mathrm{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\xi), \end{array}$$

$\phi(L)$ ia an autoregressive polynomials of order 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 and ξ_t are mutually uncorrelated at all leads and lags.
S is a matrix containing intervention variables, such as outliers, calendar effects...

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:

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

  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:

$$\begin{array}{lcll} \left[\begin{array}{c} b(L_k,\theta)\mathbf{x_t} \\ y_t \end{array}\right] & = & \vartheta_0 f_t + \gamma_t + \mathbf{S}_t\beta, & t = 1,...,n, \\ \phi(L)\Delta f_t & = & \eta_t, & \eta_t \sim \mathrm{NID}(0,\sigma^2_\eta), \\ \mathbf{D}(L)\Delta\boldsymbol{\gamma}_t & = & \boldsymbol{\delta} + \boldsymbol{\xi}_t, & \boldsymbol{\xi}_t \sim \mathrm{NID}(\mathbf{0},\mathbf{\Sigma}_\xi), \end{array}$$

The FAMIDAS collapse to MIXFAC if $K = 0$. In the application for Italian GDP: the common factor in difference is AR(2); the idiosyncratics in difference are AR(2)+drift, unless for GDP where is RW+drift . For the MIDAS we use Almon weights.

Figure: Monthly Indicators and Quarterly GDP- Italy

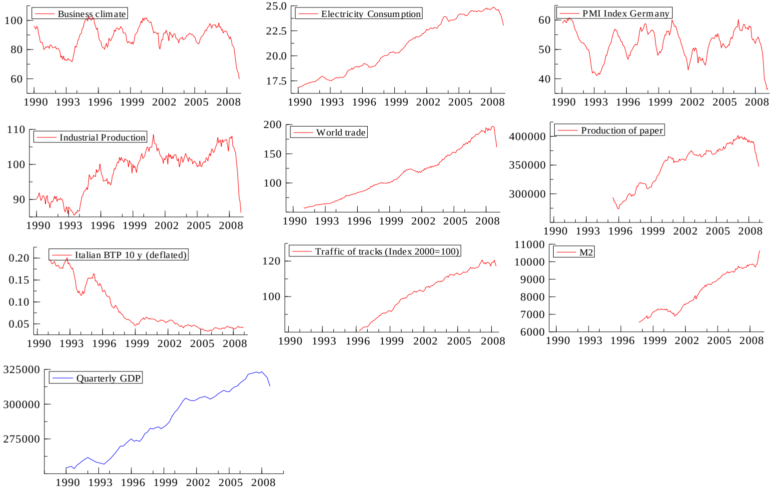


Table: Maximum likelihood estimated factor loadings (1990M1-2009M4)

	MIXFAC	MIX2FAC		FaMIDAS
		Factor 1	Factor 2	
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 **
IP	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 by Assocarta; World trade is the indicator of trade by 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.

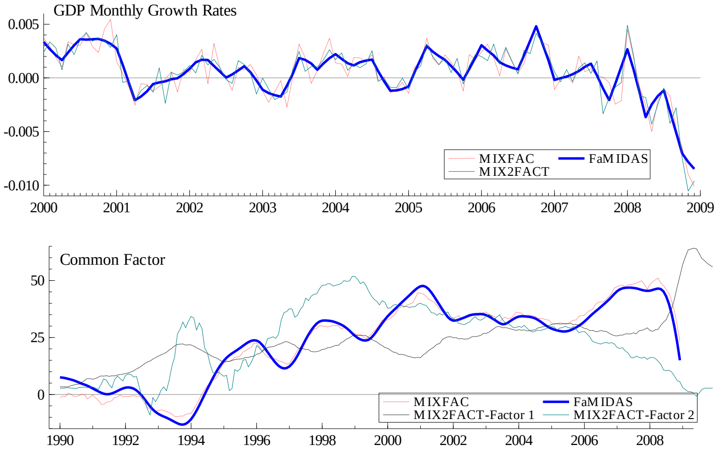


Figure: Spectral density

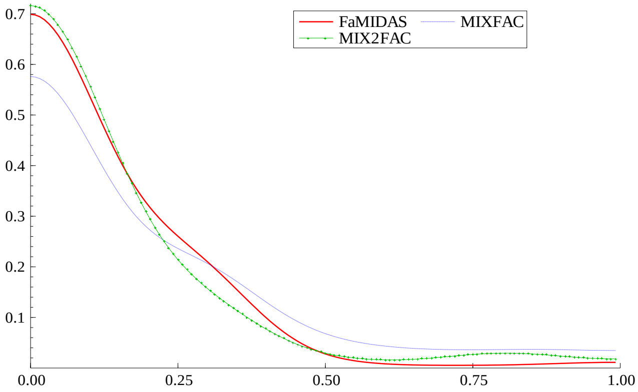
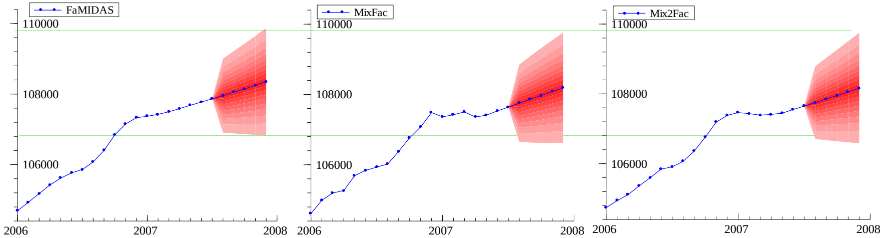


Figure: Forecasts and fan chart for the three model



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

Table: Rolling forecasting experiment for three competitor models: RMSE 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		0.40	0.43		0.35	0.43		0.34	0.38
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	0.26	0.34		0.24	0.32		0.26	0.35	
Month 2		0.31	0.36		0.30	0.35		0.32	0.38
Month 3		0.31	0.32		0.29	0.30		0.31	0.32
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	0.25	0.36		0.24	0.33		0.24	0.36	
Month 2		0.31	0.37		0.30	0.35		0.31	0.35
Month 3		0.34	0.37		0.31	0.35		0.29	0.36
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	0.28	0.35		0.26	0.34		0.30	0.36	
Month 2		0.32	0.32		0.32	0.34		0.34	0.36
Month 3		0.32	0.33		0.31	0.35		0.33	0.35
Pooling equal weights	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	0.23	0.33		0.20	0.31		0.22	0.34	
Month 2		0.29	0.33		0.28	0.33		0.30	0.35
Month 3		0.28	0.32		0.26	0.31		0.25	0.32

Note: The best values among the models (except for the pooling) are underlined. The VAR is estimated on a balanced quarterly sample. The ADL is estimated as documented by Proietti (2004)

Table: Rolling forecasting experiment for three competitor models: MAPE 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		138	145		136	162		137	78
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	116	93		120	$\underline{91}$		139	$\underline{90}$	
Month 2		$\underline{81}$	100		77	100		$\underline{73}$	99
Month 3		81	101		81	101		81	103
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	76	112		71	111		84	121	
Month 2		$\underline{81}$	$\underline{91}$		$\underline{76}$	$\underline{95}$		77	$\underline{66}$
Month 3		$\underline{79}$	95		$\underline{75}$	103		73	68
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	71	123		66	117		$\underline{69}$	137	
Month 2		90	106		85	115		80	72
Month 3		104	112		94	123		84	77
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	$\underline{69}$	$\underline{92}$		$\underline{62}$	$\underline{91}$		76	91	
Month 2		$\underline{81}$	$\underline{91}$		77	103		$\underline{73}$	68
Month 3		81	$\underline{90}$		76	$\underline{101}$		$\underline{71}$	$\underline{60}$
Pooling equal weights	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	59	103		53	101		58	111	
Month 2		75	94		71	103		67	69
Month 3		67	94		59	104		46	65

Note: The best values among the models (except for the pooling) are underlined. The VAR is estimated on a balanced quarterly sample. The ADL is estimated as documented by Proietti (2004)

Forecasting ability

- ✓ MIXFAC outperforms the other two models for nowcasting with small information set.
- ✓ MIX2FAC seems to perform best for nowcasting with complete informationset (past and current 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.

Diebold-Mariano tests

Table: Diebold-Mariano test of equal forecasting ability for horizon of previsions and month in the quarter- Rolling forecasts in a window of 5 years (2003-2007)

FaMIDAS versus Mixfac (Student t)		
	1step	2step
Month 1	2.4	1.0
Month 2	0.7	-1.9
Month 3	1.5	0.8
total	1.3	-0.5
FaMIDAS versus Mix2fac (Student t)		
	1step	2step
Month 1	1.5	-0.8
Month 2	0.6	-2.1
Month 3	-0.8	-1.9
total	0.2	-1.6

Note: Student-T are adjusted by using the Newey-West correction.

Summary

- 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.
- In our approach the MIDAS is used only to include lags of the indicators not for solving the mixed frequency issue
- At least two main advantages:
  - ✓ FaMIDAS is parsimonious and allows explicitly take into account the cross correlation between indicators and the target variable.
  - ✓ FaMIDAS mitigates the effect of data revisions in the estimates and thus on the forecasts

Conclusion

- In the empirical application the FaMIDAS produces smoother estimates for the disaggregate target variable and better forecast in a longer horizon.
- Results for the forecasting exercise provide interesting evidence:
  - The class of Multivariate mix frequency models encompass univariate models still in mixed frequency and multivariate standard models as VAR
  - The forecasting ability depends on
    - ✓ horizon of prevision: MIXFAC and MIX2FAC nowcasting; FAMIDAS for forecasting
    - ✓ information set: when complete MIX2FAC is the best
    - ✓ loss function: with MAPE the best is FaMIDAS
  - DM test confirms that differences in terms of horizon of prevision are statistically significant
  - The simplest pooling (equal weights) improves the RMSFE

Further research

- ✓ Extended the set of indicators (more financial variables, fiscal variables, services sector)
- ✓ 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.
- ✓ Exploit the use of FaMIDAS for disaggregation in sample, namely built up a monthly GDP.