LUISS Guido Carli
LIBERA UNIVERSITA' INTERNAZIONALE DEGLI STUDI SOCIALI

Luiss Lab of European Economics

LLEE Working Document no.84

FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure

Cecilia Frale, Libero Monteforte

October 2009

FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure

Cecilia Frale[1], Libero Monteforte[2] (*)

LLEE Working Document No. 84 October 2009

Abstract

The literature on short term forecasting suggests that there is room for improvement in forecast ability considering a richer dynamic structure. To address this issue, multivariate factor models are combined with MIDAS dynamics to perform short term forecasts. A dynamic factor model with mixed frequency is used, where the past observations of high frequency indicators are included following the MIDAS approach. This structure is able to represent in a richer dynamics the information content of the economic indicators and produces smoothed factors and forecasts. The short term forecasting performance of the model is evaluated against other models in a pseudo-real time experiment, also allowing for pooled forecast from factor models.

(*) Routines are coded in Ox 3.3 by Doornik (2001) and are based on the programs realized by Tommaso Proietti for the Eurostat project on EuroMIND: the Monthly Indicator of Economic Activity in the Euro Area. This paper represents the authors personal opinions and does not reflect the view of the Bank of Italy and the Italian Department of Treasury.

[1] Italian Department of Treasury; corresponding author. MEF-Ministry of the Economy and Finance-Italy, via XX Settembre 97, 00187 ROMA; E-mail: cecilia.frale@tesoro.it.

[2] Bank of Italy and Italian Department of Treasury

1 Introduction

The transmission of the recent financial crisis on the real economy was partially underestimated by a number of forecasters. Both academia and institutions are now thinking about the ability of macroeconometric models in predicting the economy and in capturing early signals of turning points. Actually the practice of short-term forecasting mainly relies on two sets of instruments: bridge models and factor models. Bridge models link timely indicators with low frequency target variables, whereas factors models extract a common component by a set (eventually large) of series [1]. In their standard formulation, bridge and factor models have shown some limitation respect to two mayor topics: the disaggregation in time and the ragged-edge data problem, which is a relevant issue for real time forecasts.

Recently we have observed a flourishing of the literature on these two approaches with extensions in different directions, among which a promising field of research is represented by mixed frequency models. The last are particularly useful to extract the information content from high frequency indicators that are used as proxy for target variables observed at lower frequency and eventually with a time lag. This is the typical requirement by economic forecasters day by day, so that the increasing interest in the literature on these models is well justified. Moreover, these models are also particularly suited as a time series disaggregation tools, given that the target variable is estimated at higher frequency.

The mixed frequency literature has initially been developed using state space factor models, estimated via the Kalman filter. Most of the applications exploit monthly series, like industrial production or confidence survey, to predict the quarterly GDP. This approach has recently been followed by Mariano and Murasawa (2003), Mittnik and Zadrozny (2004), Aruoba et al. (2007), Camacho and Perez Quiros (2008) and Frale et al. (2009). These models can also be used as a multivariate tool for time series disaggregation, as done in Frale et al. (2008), Harvey and Chung (2000), Moauro and Savio(2005).

A different approach relates to the recent literature on Mixed Data Sampling Regression Models (MIDAS) proposed by Ghysels, Santa-Clara and Valkanov (2002, 2006). MIDAS mainly differ from mixed frequency factor models as they are univariate, where lag polynomials are used to combine high frequency indicators with the low frequency target variable. There is a small but fast growing literature on MIDAS models. Most of the early applications refer to financial econometrics but there has recently been a number

[1] On the comparison on the different models for short term predictions see Barhoumi, Benk, Cristadoro, Reijer, Jakaitiene, Jelonek, Rua, Rnstler, Ruth and Nieuwenhuyze (2008).

of papers on GDP and inflation.

Clements and Galvao (2007) suggest a MIDAS to forecast monthly US quarterly macro variables or to track daily survey expectations on US macro variables (Ghysels and Wright (2006)) . Monteforte and Moretti (2008) propose a MIDAS to predict daily the monthly inflation in real time. Finally, Marcellino and Schumacher (2008) use a MIDAS to deal with an unbalanced large data-set and for estimating the monthly GDP.

This paper deals with a mixed frequency factor model and dynamic structure of the factor based on MIDAS polynomial. This feature is new in the literature and enables to exploit in a parsimonious way a larger number of lags of the high frequency indicators. This is particulary useful in forecasting as it allows to explicitly take into account the cross correlation between indicators and the target variable. Moreover, the MIDAS polynomial produces smooth factors, which is a desirable property as it imply less volatile forecasts.

In the empirical application, with Italian data, the predicting performance of the Mixed Frequency Factor Midas (in the following FaMIDAS) is compared with univariate (ADL) and multivariate (VAR) standard models and with two mixed-frequency factor models (with single and multiple factors). The results seem to suggest that the FaMIDAS prevails at larger horizons in real time forecasting. This is not surprising, as the factor produced by FaMIDAS is smooth and thus less affected by the short run variability of the data. Next Section sketches the model, while Section 3 deals with estimation and data issues. Section 4 reports the results of the forecasting exercise and Section 5 concludes.

2 The Model

Standard factor models extract a common, unobserved, component from a set of time series. In real time the series are observed with different frequency and delay therefore a mixed frequency structure avoids the problem of temporal aggregation bias. Generally factor models data-sets are unbalanced (the so called ragged-edge data) and thus a mixed frequency model works as a bridge from frequent and timely indicators toward series which are aggregated and published with higher delay. For forecast analysis, the problem of ragged-edge data calls for an efficient and explicit use of the cross correlation among variables, exploiting at best the (eventual) leading power of the timely indicators. This issue is solved in factor models including a lag (dynamic) structure, as for example imposing that the common factors follow autoregressive processes. To improve this basic framework a richer structure can be considered taking for each indicator a larger number of lags and restricting them according to a MIDAS regression. All in all, the idea is to combine the dynamic mixed frequency factor model technique with the MIDAS regression applied to the lagged indicators.

This approach might be considered as an attempt to increase the flexibility of the factor model and thus to improve its ability to reproduce the underlying structural economic model in a framework that is substantially a reduced form. As matter of fact factor models are pure statistical models, with lack of economic interpretation. Therefore, including a richer dynamic may be seen as an indirect way to capture the behavior of economic agents, as for example their expectation formation process, which might induce an effect in the correlation among time series.

In the following the two main ingredients of the model, and the way in which they are integrated, are presented.

2.1 The factor model with mixed frequency

There are many possibilities of linking a set of indicators available at high frequency to the target variable observed at lower time interval. Among them the mixed frequency factor model proposed in Frale et al.(2008) features an institutional relevance, given that it has been developed by Eurostat for EuroMIND, the Monthly INDicator of the economic activity in the Euro Area. In particular, this is a dynamic factor model that decomposes a vector of N time series, $\mathbf{y}_t$, with mixed frequency (e.g. monthly and quarterly), into one (or more) common nonstationary component, f_t, and some idiosyncratics,γ_t, specific to each series. Both the common factor and the idiosyncratics follow autoregressive standard processes as shown by the following representation:

$$\begin{array}{lcll} \mathbf{y}_t & = & \boldsymbol{\vartheta}_0 f_t + \boldsymbol{\vartheta}_1 f_{t-1} + \boldsymbol{\gamma}_t + \mathbf{S}_t \boldsymbol{\beta}, & t = 1, ..., n, \\ \phi(L)\Delta f_t & = & \eta_t, & \eta_t \sim \text{NID}(0, \sigma^2_\eta), \\ \mathbf{D}(L)\Delta\boldsymbol{\gamma}_t & = & \boldsymbol{\delta} + \boldsymbol{\eta}^*_t, & \boldsymbol{\eta}^*_t \sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_{\eta^*}), \end{array} \quad (1)$$

where $\phi(L)$ is an autoregressive polynomial of order p with stationary roots and $\mathbf{D}(L)$ is a diagonal matrix containing autoregressive polynomials of order p_i (i=1 to N) . The regression matrix $\mathbf{S}_t$ contains the values of exogenous variables that are used to incorporate calendar effects (trading day regressors, Easter, length of the month) and intervention variables (level shifts, additive outliers, etc.), and the elements of $\boldsymbol{\beta}$ that are used for initialisation and other fixed effects. The disturbances η_t and $\boldsymbol{\eta}^*_t$ are mutually uncorrelated at all leads and lags.

The model is cast in a linear State Space Form and, assuming that the disturbances have a Gaussian distribution, the unknown parameters are estimated by maximum likelihood, using the prediction error decomposition, performed by the Kalman filter (see the

Appendix for the details of the state-space representation).

The SSF is suitable modified to take into account the mixed frequency nature of the series. Following Harvey (1989), the state vector is augmented by an hoc cumulator function which translate the problem of aggregation in time into a problem of missing values. The cumulator is defined as the observed aggregated series at the end of the season (e.g. last month of quarter), otherwise it contains the partial cumulative sum of the disaggregated values (e.g. months) making up the aggregation interval (e.g. quarter) up to and including the current one.

Given the multivariate nature of the model and the mixed frequency constraint, the maximum likelihood estimation can be numerically complex. Therefore, the univariate filter and smoother for multivariate models proposed by Koopman and Durbin (2000) is used as it provides a very flexible and convenient device for handling high dimension and missing values. The main idea is that the multivariate vectors $\mathbf{y}_t$, $t = 1, \ldots, n$, where some elements can be missing, are stacked one on top of the other to yield a univariate time series $\{y_{t,i}, i = 1, \ldots, N, t = 1, \ldots, n\}$, whose elements are processed sequentially.

2.2 The MIDAS for the lags combination

As well known in the literature of leading indicators, the anticipating power of an economic series for any target variable is purely an empirical aspect. Even more cumbersome is the case of mixed frequency data, where the indicators are available at higher frequency respect to the target, so that neither the autocorrelation analysis could help. An efficient and suitable solution to this issue may be the application of the MIDAS models, which summarize and combine the information content of the indicators and their lags, with weights jointly estimated.

MIxed DAta Sampling models have recently encountered a considerable success due to their simplicity and good performance in empirical applications. To introduce them, as in the seminal paper by Ghysels et al.(2004), suppose Y_t is a time series variable observed at a certain fixed frequency and let X^m be and indicator variable sampled m times faster. A MIDAS regression takes the form:

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

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.

Mainly the MIDAS structure involves two elements: the reconciliation of different frequency and the use of lagged values of the indicators.

In our application the MIDAS is applied only to exploit efficiently the lag structure of indicators, whereas the time aggregation problem is solved inside the factor model as shown in section 2(namely this is MIDAS with m=1). This allows better interpretation of the cyclical pattern of the economic indicators and comparability with benchmark dynamic models.

As the weights structure is concerned, two main possibilities have been proposed in the literature. First, a parametrization that refers to Almon lags:

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

Second, 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 $\Gamma(a) = \int_0^\infty e^(-x)x^{a-1}dx$.

There is not a clear a priori reason for preferring one parametrization respect to the other, and the choice should clearly be contingent to the research problem under analysis. As rule of thumb it could be mentioned that the Beta function, given its flexibility, seems more suitable when the number of lags considered is large whereas the simplicity of the Almon weights might be preferable in the case of small number of time lags involved.

Looking at the recent literature, Marcellino and Schumacher(2008) used the Almon weights for the estimation of GDP in real time, whereas Moretti and Monteforte (2008) found the Beta transformation more appropriate for the estimation of inflation which involves daily data and more than 20 lags.

2.3 The FaMIDAS

This section presents how combine the dynamic factor model with mixed frequency and the MIDAS structure of lags described in the previous section.

Starting from model 1 let us partitioning the set of time series, $\mathbf{y}_t$, into two groups, $\mathbf{y}_t = [\mathbf{y}'_{1,t}, \mathbf{y}'_{2,t}]'$, where the second block represent the target variable available at lower frequency. Now let introduce the MIDAS structure for the high frequency indicators so that $\mathbf{y}'_{1,t} = [b(L_k,\theta)\mathbf{x_t}]'$.

The FaMIDAS results by the following equations:

$$
\begin{array}{lll}
\begin{bmatrix} b(L_k,\theta)\mathbf{x_t} \\ \mathbf{y}_{2,t} \end{bmatrix} & = \boldsymbol{\vartheta}_0 f_t + \boldsymbol{\gamma}_t + \mathbf{S}_t\boldsymbol{\beta}, & t = 1,...,n, \\
\phi(L)\Delta f_t & = \eta_t, & \eta_t \sim \text{NID}(0,\sigma^2_\eta), \\
\mathbf{D}(L)\Delta\boldsymbol{\gamma}_t & = \boldsymbol{\delta} + \boldsymbol{\eta}^*_t, & \boldsymbol{\eta}^*_t \sim \text{NID}(\mathbf{0},\boldsymbol{\Sigma}_{\eta^*}),
\end{array}
\tag{2}
$$

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

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

For identification reasons, the weights sum up to 1 so that their size is fully comparable. In particular, the relative size of the weights associated to the lags allows inferring the leading or coincident power of the series. For example, a decreasing pattern in the weights vector, associated with the lags $(x_t, x_{t-1}, x_{t-2}...x_{t-k})'$, might be considered as an indication of a coincident relation between the indicator and the target, while an increasing pattern weights suggests a leading relationship.

As the maximum lag length is concerned, the target horizon of forecasting and the economic meaning of the series could suggest the appropriate number.

3 The Empirical Application

In this section we present the results as for the application of model 2 to the estimation and forecasting of the Italian GDP at the monthly level.

We exploit the information of the most relevant monthly economic indicators, available earlier than the official statistics, to disaggregate, nowcast and forecast the quarterly GDP. Therefore an estimate of the unobserved monthly GDP is obtained, both for the past (a monthly indicator of the known quarterly GDP) and for the future. It is worth to note that the model is such that the monthly indicator is fully consistent with the quarterly data, in terms of time aggregation. Thus we obtain an indicator that can be used both in sample as monthly measure of GDP and out of sample as leading indicator.

In particular, when the main goal is forecasting, the leading power of the indicators could be efficiently exploited by including in the model their lags. Instead of choosing a priori which and how many lags to include, the MIDAS model is chosen to summarize and combine the information of the dynamics.

A two steps estimation procedure is proposed for the faMIDAS. First, the anticipating power of the economic indicators is used efficiently by means of the recent class of Mixed Data Sampling Regression Models (MIDAS). Second, the results are plugged in a dynamic factor model with mixed frequency (as in Frale et al. (2008)), which links the target variable (e.g. quarterly GDP) to composite series available at higher frequency (e.g. MIDAS linear combination of lags of the monthly economic indicators). Starting from a trial, the procedure is run iteratively so that the weights in the MIDAS maximize the Likelihood function associated with the factor model.

The GDP is estimated directly, leaving the bottom-up approach (estimation by aggregation of sectorial value added and $\backslash$ or components of demand) for future research. Although the model is specified in level in order to easily deal with the time constraint, the results and the forecasting experiment are presented in growth rate, which is the reference measure for both policy makers and academics.

As the variable selection is concerned, a large set of indicators is considered, with series referring to different aspects of the economy. These are mainly official hard data, such as industrial production, world trade, production of paper; survey data, such as climate, expectation and PMI; financial data, such as spread and money (M2) and private sources such as electricity consumption, traffic of trucks.

For the model selection process we follow the traditional approach in the literature, based on: statistical significance of the indicators, BIC and Akaike criteria for lag selection.

After some empirical robustness checks, the sample 1990M1-2009M4 has turned out to be the best trade-off among representativeness of the sample size, the availability of long time series and the quality of the indicators.

Three benchmark models have been selected. The first model (MIXFAC) is a dynamic single factor model as in Frale et al. (2008) with 4 indicators: Electricity consumption, Industrial production, German PMI, Business climate, and this is intended as the baseline model.

The second model (MIX2FAC) is a 2 factors model, as in Frale et al. (2009) including more indicators: Industrial production of paper, world trade, Treasury Italian yields (10Y), money supply, motorway flows (of trucks). All indicators are shown in figure 1.

The third model is a factor-MIDAS model (faMIDAS) with the same indicators of MIXFAC but MIDAS combination of their lags up to 4. Alternative lags length have been evaluated accordingly to the horizon of the target forecast (maximum 6 months ahead) and the economic meaning of the indicators.

Summing up, MIXFAC is the baseline model, involving both survey and national account data. In MIX2FAC more soft indicators are included and the second factor allow capturing financial swings, as it comes up ex-post. Finally using faMIDAS it is possible to consider up to four lags of the economic indicators of MIXFAC.

The estimated maximum likelihood parameters are listed in table 1, whereas the estimated weights for the MIDAS model are reported in figure 2. In particular, the analysis of the MIDAS scores reveals an increasing parabolic pattern up to the second lag and then reversing. The series of industrial production has the sharper weights path, which is coherent with the general sense of relevance of this series in mirroring the pattern of GDP.

In addition, figure 3 shows the estimated GDP in monthly growth rates and the common factors for the three models. It is clearly visible that the MIDAS estimation produces smoother forecast, which is from one side reasonable considering that it makes a sum over time lags and from the other a good desirable property. Moreover the confidence bands of the forecasts, shown as fan charts in figure 4, reveal that incertitude of the prevision is smaller in the MIDAS model than in the other mix-frequency formalizations. The following figure **??** gives a more detailed insight of the forecast precision, showing the standard error of the forecasts for the three models versus a standard ADL model. It is clear that the classical approach is outperformed by the mix frequency parametrization, especially in short term horizon. The MIXFAC model seems preferable for short term forecast (up to 1 quarter-ahead) , the MIDAS for longer prevision (up to 2 quarters-ahead), and MIX2FAC in a sort of intermediate case between those two. Moreover, the inspection of the spectral density of the estimated monthly GDP for the MIDAS and MIXFAC, shown in figure 5, suggests that the MIDAS structure is able to capture low frequencies fluctuations and therefore might perform better for longer horizons forecasts.

As matter of fact the forecasting performance analysis of the three models requires a rolling (or recursive) experiment, which is accomplished in the next section.

4 Forecasting evaluation

In this section the three models under analysis are compared respect to their forecasting ability, with a rolling experiment in a window of the latest 5,4,3 years up to the end of 2007 [2]. The rolling exercise is made in real time, so as to mimic the delay of different indicators, which has been proved to be relevant for the correct assessment of the best model. Therefore the forecasting evaluation is made with distinction of the month of the prevision inside the quarter (e.g. first month, second or third), which corresponds to a different information set. It is worth to stress that the Kalman filter is particularly suitable for this issue given that it solves endogenously the problem of the unbalanced

[2]We prefer to take the year 2008 outside the experiment to avoid that the exceptional conditions of the economic crisis invalidates the conclusion. In addition, at time of writing data from 2008 upwards are still preliminary and subject to revision.

sample produced by the different delay of publication of the monthly indicators. It is interesting to find, looking at the results reported in table 2, that MIXFAC outperforms the other two models for short term forecasts with small information sample. On the contrary, MIX2FAC seems to perform best for nowcasting with medium information set (second month of the quarter) while the FaMIDAS makes the lowest forecasting error for the prevision 1 quarter-ahead. This last empirical evidence might reinforce the idea that the FaMIDAS exploit efficiently the correlation of the lag structure of the indicators with the target variable.

Since the seminal paper by Bates and Granger(1969), it is well know that combining different models can get to a smaller forecast error that selecting a single specification. Stock and Watson (1999) presented a detail analysis of the forecast improvement by using pooling forecast for a huge set of macroeconomic indicators in the US, whereas Marcellino (2004) proposed a similar exercise for the Euro Area. Moreover Kuzin et al (2009) compared the ability of forecasting the Germany GDP over many specifications, including MIDAS and factor models, concluding that the empirical evidence provides support for pooling rather than single models. The general idea is that the combination of different specifications, by averaging, mitigate the measurement error of each specific model. Therefore, the pooling forecast is particularly suitable when the combined models show a significant heterogeneity.

The application presented above matches this requirement, given that the models differ in terms of components (number of factors and lags), as well as for the best horizon of forecast. In the bottom panel of table 2 we report the RMSFE in real time for the pooled model with equal weights, which show that the combination of the three models, the MIXFAC, MIX2FAC and MIDAS is preferred to each of them singularly. In fact the forecasts produced by the pooling of different models overpass the single models, with only few exceptions for which anyway the RMSFE is very close to the smallest. This result is coherent with the theory of pooling forecasts and thus provide evidence on the robustness of the empirical application. Although it has been shown (see for example Drechsel and Maurin (2008)) that the weighting system might influence the results, we leave this issue for future research.

5 Conclusions

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. A factor model with mix frequency data is proposed, where some relevant economic indicators are used to disaggregate and predict a target variable which is observed with smaller time frequency. In order to efficiently capture the leading power of such indicators, a MIDAS structure, based on lags of the same indicators, is specified in the common factor. In the empirical application the model, that we call FaMIDAS, is tested versus standard univariate and multivariate models, including mixed frequency factor models. Overall the MIDAS produces smoother estimates for the disaggregate target variable and better forecast in a longer horizon. In order to reduce forecast variability pooling methods are proposed. Results for this application confirm some evidence in the literature that pooling forecasts are more stable than previsions from single models.

Appendix: The State space representation and temporal aggregation

Model (1) described in the previous section could be easily written in the state space form (SSF) as follow. Consider the standard way to recast in SSF a general AR(p) process $\phi(L)\Delta\chi_t = \eta_t$ with $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$:

$$\chi_t = \mathbf{e}'_{1,p+1}\boldsymbol{\alpha}_t, \quad \boldsymbol{\alpha}_t = \mathbf{T}_\chi \boldsymbol{\alpha}_{t-1} + \mathbf{H}\eta_t,$$

where

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \chi_t \\ \chi_t^* \end{bmatrix}, \quad \mathbf{T}_\chi = \begin{bmatrix} 1 & \mathbf{e}'_{1p}\mathbf{T}_\phi \\ 0 & \mathbf{T}_\phi \end{bmatrix}, \mathbf{T}_\phi = \begin{bmatrix} \phi_1 & \\ \vdots & \mathbf{I}_{p-1} \\ \phi_{p-1} & \\ \phi_p & \mathbf{0}' \end{bmatrix}.$$

and $\chi_t^* = \mathbf{T}_\phi \chi_{t-1}^* + \mathbf{e}_{1p}\eta_t$, $\mathbf{H} = [1, \mathbf{e}'_{1,p}]'$,$\mathbf{e}_{1p} = [1, 0, \dots, 0]'$.

This representation holds for the common factor and for each idiosyncratics, therefore the model could be obtained applying the previous representation to all components and combining them stacking the single elements in the state vector $[\alpha_t = \alpha'_{f,t}, \alpha'_{\gamma_1,t}, \dots, \alpha'_{\gamma_N,t}]'$, in the vector of errors $\boldsymbol{\epsilon}_t = [\eta_t, \eta^*_{1,t}, \dots, \eta^*_{N,t}]'$ and in the system matrices of the measurement equation:

$$\begin{aligned} &\mathbf{Z} = \left[\boldsymbol{\theta}_0, \vdots\, \boldsymbol{\theta}_1 \ \vdots\, \mathbf{0} \,\vdots\, \text{diag}(\mathbf{e}'_{p_1}, \dots, \mathbf{e}'_{p_N})\right], \qquad \mathbf{T} = \text{diag}(\mathbf{T}_f, \mathbf{T}_{\gamma_1}, \dots, \mathbf{T}_{\gamma_N}), \\ &\mathbf{H} = \text{diag}(\mathbf{H}_f, \mathbf{H}_{\gamma_1}, \dots, \mathbf{H}_{\gamma_N}). \end{aligned} \tag{3}$$

Finally, the SSF of the complete model results:

$$\mathbf{y}_t = \mathbf{Z}\boldsymbol{\alpha}_t + \mathbf{S}_t\boldsymbol{\beta}, \quad \boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{W}\boldsymbol{b} + \mathbf{H}\boldsymbol{\epsilon}_t, \tag{4}$$

where $\mathbf{W}$ is a time invariant matrix that selects the drift δ_i for the appropriate state element of the idiosyncratic component.

The temporal aggregation problem is solved following the strategy proposed by Harvey (1989). The set of indicators, $\mathbf{y}_t$, is partitioned into two groups, $\mathbf{y}_t = [\mathbf{y}'_{1,t}, \mathbf{y}'_{2,t}]'$, where the second block gathers the flows that are subject to temporal aggregation. In addition, the high frequency indicators are summarized in a MIDAS structure, such as $\mathbf{y}'_{1,t} = [b(L_k, \theta)\mathbf{x_t}]'$.

An *ad hoc* cumulator variable, $\mathbf{y}^c_{2,t}$, is defined so that it coincides with the (observed) aggregated series at the end of the larger interval (e.g. quarter), otherwise it contains the partial cumulative value of the aggregate in the seasons (e.g. months), as follow:

$$\mathbf{y}^c_{2,t} = \psi_t \mathbf{y}^c_{2,t-1} + \mathbf{y}_{2,t}, \qquad \psi_t = \begin{cases} 0 & t = \delta(\tau - 1) + 1, \quad \tau = 1, \dots, [n/\delta] \\ 1 & \text{otherwise} , \end{cases}$$

The cumulator is used to replace the second block of the measurement equation and to augment the state equation as follow:

$$\boldsymbol{\alpha}^*_t = \begin{bmatrix} \boldsymbol{\alpha}_t \\ \mathbf{y}^c_{2,t} \end{bmatrix}, \quad \mathbf{y}^\dagger_t = \begin{bmatrix} b(L_k, \theta)\mathbf{x_t} \\ \mathbf{y}^c_{2,t} \end{bmatrix}$$

The final measurement and transition equation are therefore:

$$\mathbf{y}^\dagger_t = \mathbf{Z}^* \boldsymbol{\alpha}^*_t + \mathbf{S}_t \boldsymbol{\beta}, \quad \boldsymbol{\alpha}^*_t = \mathbf{T}^* \boldsymbol{\alpha}^*_{t-1} + \mathbf{W}^* \boldsymbol{\beta} + \mathbf{H}^* \boldsymbol{\epsilon}_t, \tag{5}$$

with system matrices:

$$\mathbf{Z}^* = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_2} \end{bmatrix}, \quad \mathbf{T}^* = \begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{Z}_2\mathbf{T} & \psi_t \mathbf{I} \end{bmatrix}, \quad \mathbf{W}^* = \begin{bmatrix} \mathbf{W} \\ \mathbf{Z}_2\mathbf{W} + \mathbf{S}_2 \end{bmatrix}, \quad \mathbf{H}^* = \begin{bmatrix} \mathbf{I} \\ \mathbf{Z}_2 \end{bmatrix} \mathbf{H}. \tag{6}$$

References

Aruoba, S. B., F. X. Diebold and C. Scotti, (2007), ”Real-time Measurement of Business Conditions,” International Finance Discussion Papers, Board of Governors of the Federal Reserve System (U.S.), No. 901.

Bańbura M. and Rünstler (2007), A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP, European Central Bank working paper no. 751.

Barhoumi, K., Benk, S., Cristadoro, R., Reijer, A. D., Jakaitiene, A., Jelonek, P., Rua, A., Rnstler, G., Ruth, K. and C. V. Nieuwenhuyze (2008). Short-term forecasting of GDP using large monthly datasets: a pseudo real-time forecast evaluation exercise, ECB Occasional papers No. 84.

Bates, J.M., Granger, C.W.J. (1969), ”The combination of forecasts”, Operations Research Quarterly, 20, 415-468.

Camacho M., Perez-Quiros G. Introducing the Euro-Sting: short term indicator of the Euro Area growth, Banco de España working paper 0807, 2008.

Clements, M. P. and A. B. Galvo, (2007), ”Macroeconomic Forecast ing with Mixed Frequency Data: Forecasting US Output Growth and In?ation”, Queen Mary, University of London, Working Paper No. 616.

Doornik, J.A. (2001). *Ox 3.0 - An Object-Oriented Matrix Programming Language*, Timberlake Consultants Ltd: London.

Drechsel, K., Maurin, L.(2008), ”Flow of Conjunctural Information and Forecast of the Euro Area Economic Activity”, ECB Working Paper No. 925.

Durbin, J. and Koopman, S.J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, UK.

Frale C., Marcellino M., Mazzi G. L., Proietti T.(2008), A Monthly Indicator for Euro Area GDP, CEPR Discussion Paper 7007.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic Factor Model: identifcation and estimation, *Review of Economics and Statistics*, 82, 540-554

Geweke J. (1977). The dynamic factor analysis of economic time series models. In *Latent Variables in Socio-Economic Models*, Aigner DJ, Goldberger AS (eds); North Holland: New York.

Giannone, D., Reichlin, L., and Small, D.(2005) Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases, Finance and Economics Discussion Series 42.

Ghysels, E., P. Santa-Clara and R. Valkanov (2002), ”The MIDAS Touch: Mixed Data Sampling Regression Models,” UNC and UCLA Working Papers.

Ghysels, E., P. Santa-Clara and R. Valkanov (2006), ”Predicting Volatility: Getting the-Most Out of Return Data Sampled at Different Frequencies”, Journal of Econometrics, No. 131(1-2).

Ghysels, E. and J. H. Wright, (2008), ”Forecasting Professional Forecasters,” Journal of Business, Economics and Statistics, forthcoming.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, Cambridge University Press.

Harvey, A.C. and Chung, C.H. (2000) Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistics Society*, Series A, 163, 303-339.

Harvey,A.C.and Chung,C.H.(2000)Estimating the underlying change in unemployment in the UK.Journal of the Royal Statistics Society,Series A,163,303-339.

Harvey, A.C., and Pierse R.G. (1984). Estimating Missing Observations in Economic Time Series. *Journal of the American Statistical Association*, 79, 125-131.

Koopman, S.J., and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models, *Journal of Time Series Analysis*, 21, 281–296.

Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series, *Journal of Business and Economic Statistics*, 1, 2, pp. 169-173.

Kuzin, V., Marcellino, M.,Schumacher C. (2009), Pooling versus model selection for nowcasting with many predictors: An application to German GDP, CEPR Discussion Paper 7197.

Marcellino, M. (2007), Pooling-based data Interpolation and Backdating, CEPR Discussion Paper 5295.

Marcellino, M. (2004), Forecast Pooling for European Macroeconomic Variables,Oxford Bulletin of Economics and Statistics, vol. 66, no. 1, pp. 91-112.

Marcellino, M. and C. Schumacher, (2008), ”Factor-MIDAS for Now- and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP”,CEPR working paper, No. 6708.

Mariano, R.S. and Y. Murasawa (2003), ”A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series,” Journal of Applied Econometrics, No. 18.

Mitchell, J., Smith, R.J., Weale, M.R., Wright, S. and Salazar, E.L. (2004), An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth. Discussion Paper 127 (revised) National Institute of Economic and Social Research.

Mittnik, S. and P. Zadrozny, (2004), ”Forecasting Quarterly German GP at MOnthly Interval Using Monthly IFO Business Condition Data,” CESifo working paper, No. 1203.

Moauro F. and Savio G.(2005).”Temporal Disaggregation Using Multivariate Structural Time Series Models”, Econometrics Journal,8,214-234.

Monteforte, L. and G. Moretti (2008), ”Real Time Forecasts of Inflation: The Role of Financial Variables”, Available at SSRN: http://ssrn.com/abstract=1263272.

Proietti, T. (2004). Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited. *Econometrics Journal*, 9, 357-372.

Proietti, T. (2006). On the estimation of nonlinearly aggregated mixed models. *Journal of Computational and Graphical Statistics*, Vol. 15, 1–21.

Stock, J.H., and Watson M.W. (1991). A probability model of the coincident economic indicators. In *Leading Economic Indicators*, Lahiri K, Moore GH (eds), Cambridge University Press, New York.

Stock, J.H., and Watson M.W. (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series", in Engle, R. and White, R. (eds), *Cointegration, causality and forecasting: A festschrift in honor of Clive W.J. Granger*, Oxford: Oxford University Press, 1-44.

Table 1: Maximum likelihood estimated factor loadings ϑ-(1990M1-2009M4)

	MIXFAC	**MIX2FAC**		**FaMIDAS**
		Factor 1	Factor 2	
Business Climate	0.44 **	-0.61 **	-0.02	0.09 **
Electricity	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(-1)	-0.22			
IP(-1)	0.67 **			
IP paper		-0.14 **	0.03	
World trade (CPB)		-0.74 **	0.17	
Italian BTP 10y		-0.03	-0.37**	
M2		0.24 **	-0.02	
Traffic of 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 2: Rolling forecasting experiment for three competitor models: 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 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}
	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: The best values among the models (except for the pooling) are underlined. The VAR is estimated on a balanced quarterly sample.

Figure 1: Monthly Indicators and Quarterly GDP- Italy

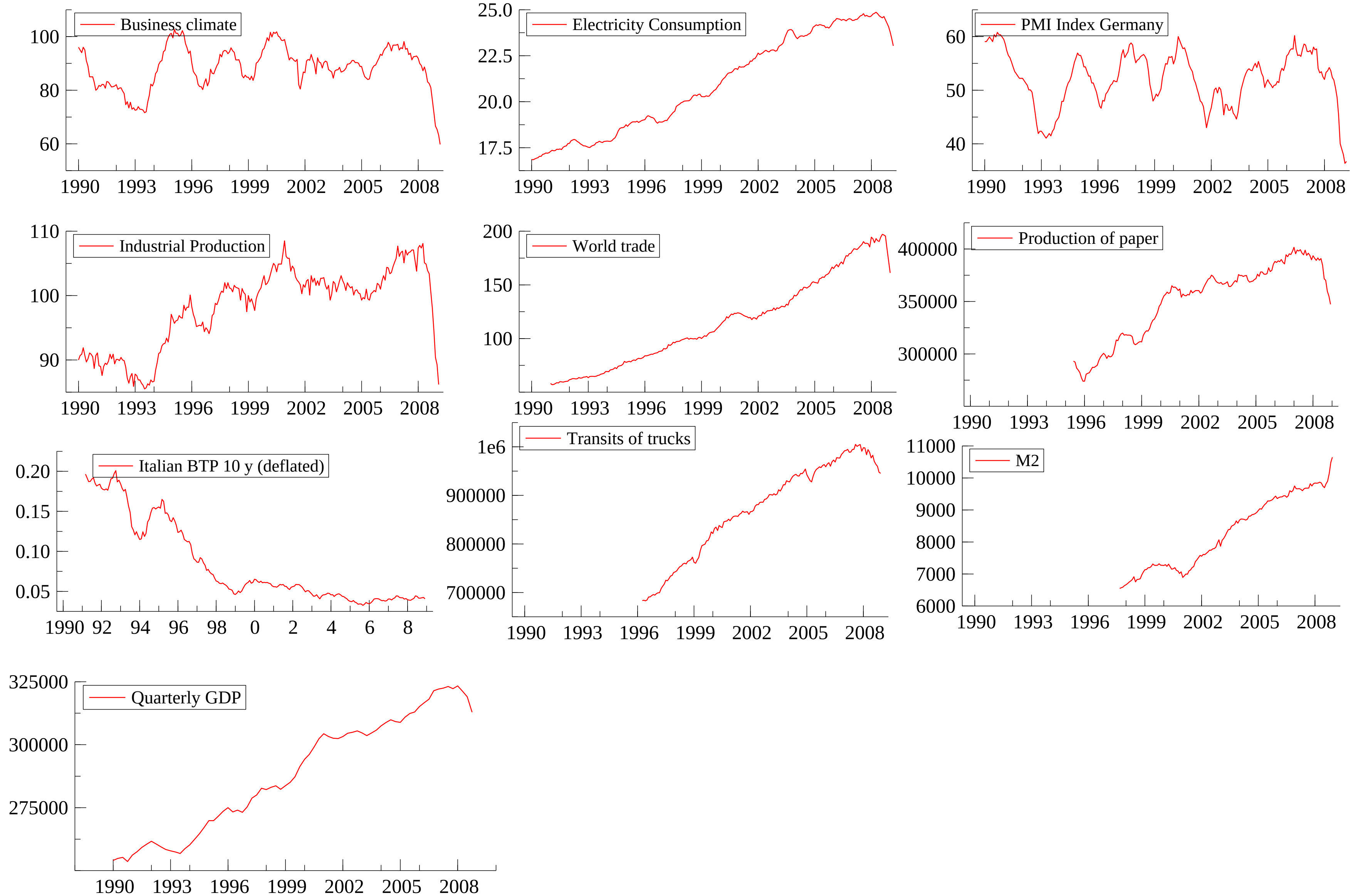


Figure 2: FaMIDAS estimated weights

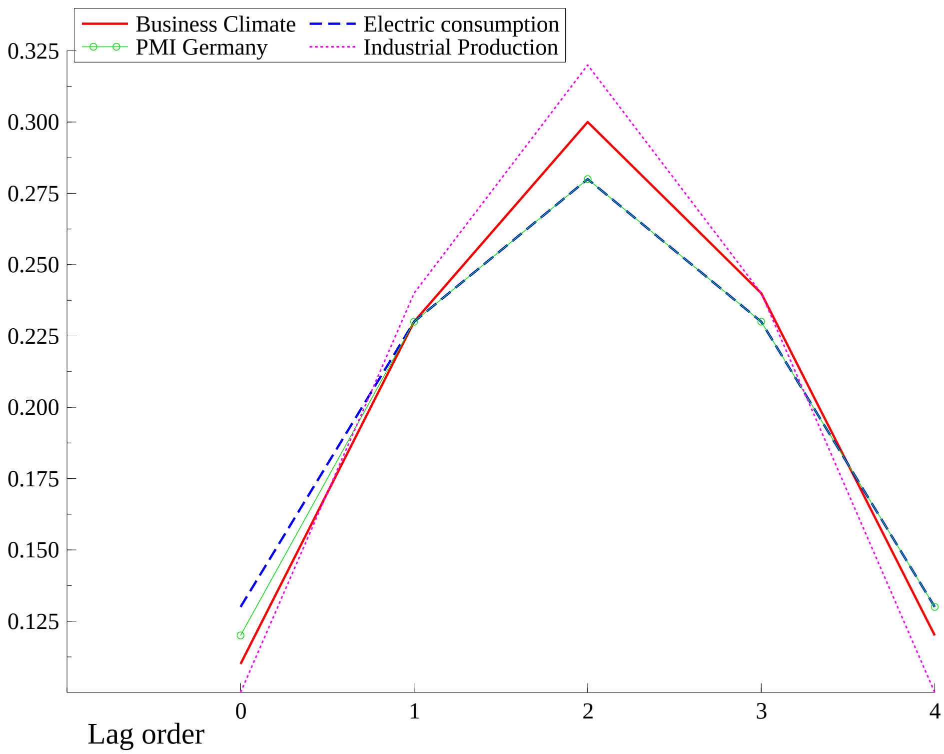


Figure 3: Estimated Monthly GDP (growth rate) and common factors for the three models.

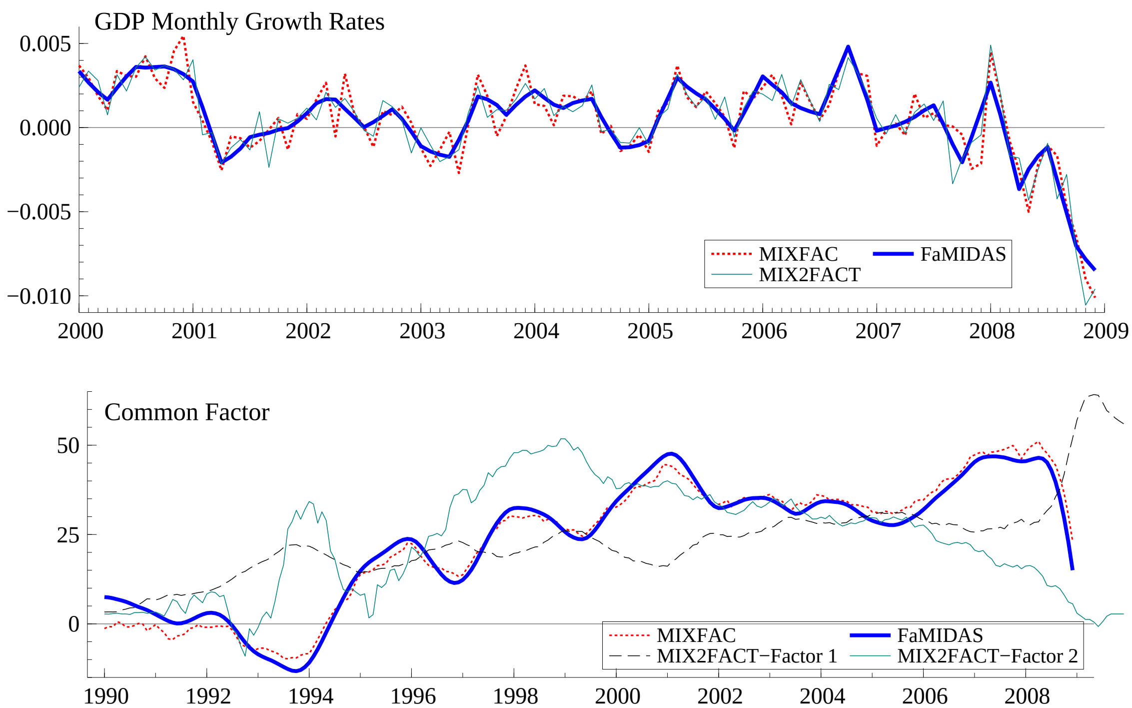
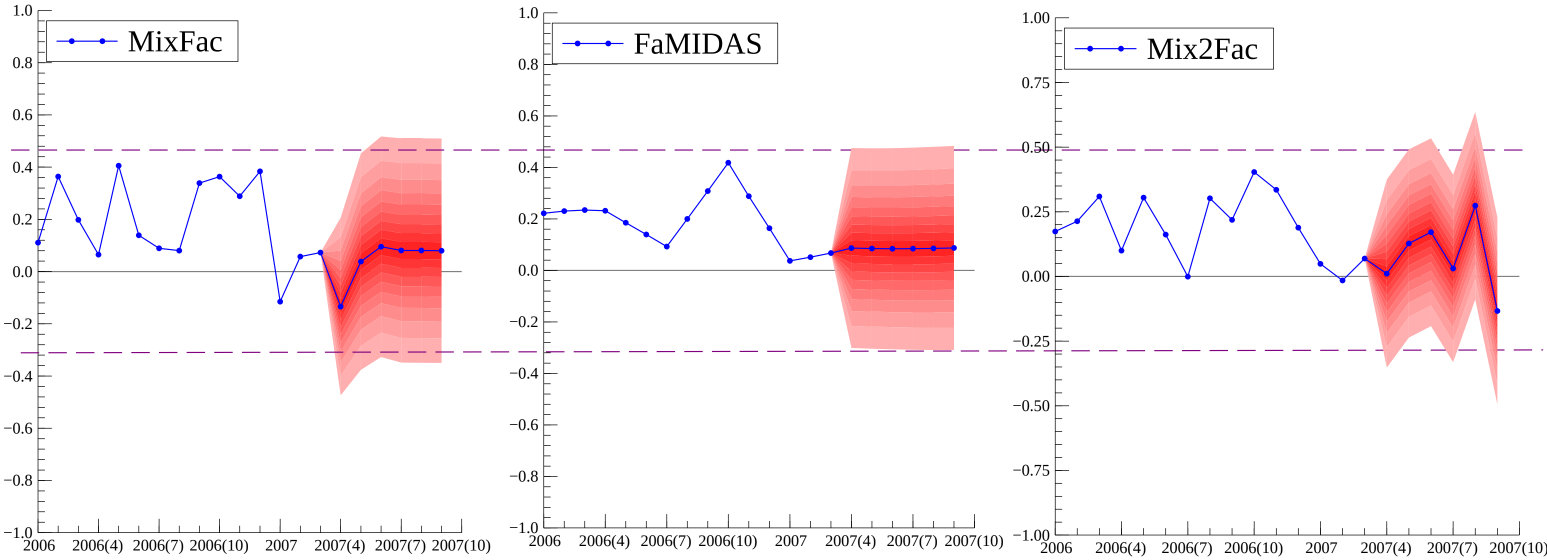


Figure 4: Forecasts and fan chart for the three model



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

Figure 5: Spectral Density of the Monthly GDP (in monthly growth rate)