Ministry of Economy and Finance

Department of the Treasury

Analisi e Programmazione
Economico Finanziaria

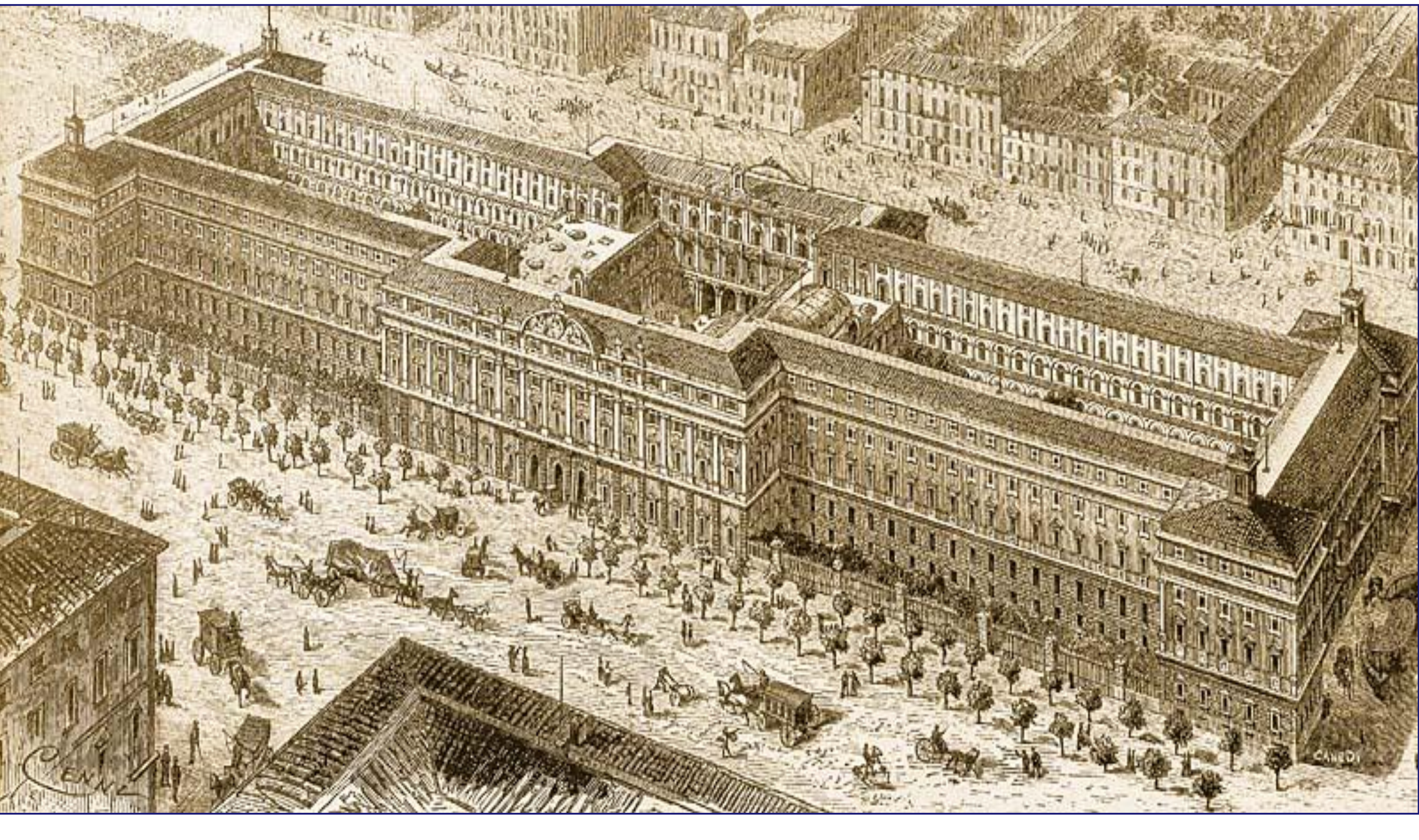
Working Papers

N°6 - March 2011

ISSN 1972-411X

Real time forecasts of inflation: the role of financial variables

Libero Monteforte, Gianluca Moretti



Working Papers

The working paper series promotes the dissemination of economic research produced in the Department of the Treasury (DT) of the Italian Ministry of Economy and Finance (MEF) or presented by external economists on the occasion of seminars organised by MEF on topics of institutional interest to the DT, with the aim of stimulating comments and suggestions.
The views expressed in the working papers are those of the authors and do not necessarily reflect those of the MEF and the DT.

Real time forecasts of inflation: the role of financial variables[1]

Libero Monteforte (*),Gianluca Moretti (**)

Abstract

We present a mixed-frequency model for daily forecasts of euro area inflation. The model combines a monthly index of core inflation with daily data from financial markets; estimates are carried out with the MIDAS regression approach. The forecasting ability of the model in real time is compared with that of standard VARs and of daily quotes of economic derivatives on euro area inflation. We find that the inclusion of daily variables helps to reduce forecast errors with respect to models that consider only monthly variables. The mixed-frequency model also displays superior predictive performance with respect to forecasts solely based on economic derivatives.

JEL Classification: C13, C51, C53, E37, G19.
Keywords: forecasting inflation, real-time forecasts, dynamic factor models, MIDAS regression, economic derivatives.

(*) Banca d'Italia and Ministero dell'Economia e delle Finanze.

(**) UBS Global Asset Management.

[1] This paper was written while Gianluca Moretti was working at the Bank of Italy. The views expressed in this article do not necessarily reflect the views of any institution with which the authors are affiliated. We are indebted to Eric Ghysels, Marco Lippi, Alessio Sancetta and Paolo Zaffaroni for useful comments. We thank partecipants of the Society for Financial Econometrics conference (SoFiE 2008), 5th Eurostat Colloquium on Modern Tools for Business Cycle Analysis, 2nd Workshop in Computational and Financial Econometrics and 20th $(EC)^2$ Conference; in particular Giovanni Barone Adesi, Luc Bauwens, Tim Bollerslev, Frank Diebold, Andrea Ghiringhelli, Menelaos Karanasos, Enno Mammen and Christian Schumacher for useful discussions. We are also grateful to Fabio Busetti for helpful comments at various stages of the work. Usual disclaimers apply.

CONTENTS

1 INTRODUCTION

Forecasting inflation as frequently as possible has become significantly important for both institutional and private operators in recent years. On the one hand, a timely update of the macroeconomic projections is essential for conducting a modern monetary policy based on the market expectations (see Woodford (2003)); on the other hand, market participants tend to update their expectations continuously as new information is released, and to exploit this information to modify their investment strategies.[1] Since inflation data are available at monthly frequency, the common approach to forecasting inflation is to construct models based on monthly variables that are correlated with future inflation. Forecasts from these models can be quite accurate but not very timely, since by construction, they do not account for any important information that might become available within the month. One solution to this problem is to look at financial indicators, such as movements in the yield curve or interest rate spreads, that are available on a daily basis and that can provide some timely information about changes in inflation expectations. An alternative solution, which we present in this paper, is to construct an indicator of monthly inflation using both monthly and daily data. More precisely, we propose a mixed-frequency data model that combines a monthly core inflation index, constructed using a generalized dynamic factor model, with daily prices of commodities and financial assets.

Factor models have become in recent years a very popular approach to construct economic indicators since they allow us to handle the information contained in a large number of variables in a parsimonious way (see for instance Stock and Watson (2006)). In particular, the generalized dynamic factor (GDF) models, proposed by Forni et al. (2002), can explain most of the variability of the data at low and medium frequency by extracting a few common components from a large set observable variables. This approach has been recently used, for instance, to construct measures of core inflation for the euro area by Cristadoro et al. (2005, 2008). Despite their popularity, large-scale factor models present a few drawbacks when they are used to forecast in a real time macroeconomic variables especially at short horizons. First, because they are designed to capture the slow-moving component of inflation, they have their best forecast accuracy at medium and long-term

[1] Bernanke and Kuttner (2005) study the reaction of asset prices to monetary policy shocks, Andersen et al. (2003) show that U.S. dollar exchange-rate quotations are linked to economic fundamentals.

horizons. Moreover, since economic data are released with different lags, factor models are characterized by an unbalanced end of sample (ragged-edge). This problem, well known in the factor model literature, can have considerable effects on the forecasting properties of the models, especially in the short-term (see Marcellino and Schumacher (2010)).

In this paper, we propose to tackle this problem using a mixed frequency data models where a factor model indicator, estimated with a balanced dataset, is combined with relevant explanatory variables sampled at a higher frequency. For the case of euro area inflation we focus on daily prices of relevant commodities and financial assets. The choice of these variables lies in their ability to capture changes in market expectations as well as movements in the most volatile components on inflation, specifically energy and food, which are highly correlated with the short-run developments of the overall index. Furthermore, they can also anticipate the effects of changes in indirect taxes or other government measures on inflation that can not be easily captured in a pure backward-looking model. In this respect, our approach can be seen as an efficient way of exploiting the predictive content of financial data without altering the temporal structure of the data.

The class of mixed-frequency models we consider is the MIxed DAta Sampling regression models (MIDAS), proposed by Ghysels et al. (2004, 2006).[2] Most of the early applications of MIDAS models were on financial data but more recently there have been a few applications to macroeconomic variables. Clements and Galvão (2008) and Andreou, Ghysels and Kourtellos (2010) use a MIDAS model to forecast US quarterly macro variables, Ghysels and Wright (2009) to track daily survey expectations of US macro variables and Marcellino and Schumacher (2010) to produce monthly estimate of German GDP. The use of asset prices to forecast macro variables is also not new in the literature. In an extensive review Stock and Watson (2003) show that financial variables have a statistically significant marginal predictive content for macroeconomic variables. However, this predictive content can not be exploited reliably and it is often very unstable over time. In this respect, our mixed-frequency approach can be an effective way to exploit the information coming from financial markets. We show this by running two real time forecasting exercises. In the first we assess the accuracy of our mixed-

[2] Another class of mixed-frequency models, proposed by Mittnik and Zadrozny (2004) and more recently by Aruoba et al. (2009), is based on a state-space representation where they use the Kalman filter to construct high frequency unobserved indicators for the low-frequency variable.

frequency indicator in forecasting euro area inflation relative to standard monthly models and market expectations. We show that predictions from our mixed-frequency model outperform those of the standard benchmark models based only on monthly variables confirming that the inclusion of the daily variables helps to reduce the forecasting errors. In the second exercise, we fully exploit the high-frequency structure of our models and run a daily forecasting exercise: we compare predictions of our model with market expectations, extracted from the quotes of euro area HICP future contracts. Daily forecasts of our models are not only more accurate than those extracted from the daily quotes of euro area HICP future contracts but also less volatile.

The paper is organized as follows. Section 2 reviews the link between financial variables and inflation and describes the economic derivatives we used in our analysis. Section 3 describes our mixed-frequency models in details. In Section 4 we assess the forecasting ability of our models compared with VAR models and market daily expectations. Section 5 concludes.

2 ROLE OF FINANCIAL VARIABLES

If keeping track of persistent movements in inflation is crucial to the pursuit of price stability, it is also important for policy makers and financial operators to monitor the evolution of current inflation in real time. This is usually done by looking at the information embodied in high-frequency variables, such as the yield curve, bond yields and quotes on HICP economic derivatives. Not only are these variables forward looking and observed in real time, but also have potentially useful information about inflation expectations. Moreover, unlike macroeconomic data, they are not subject to any revisions. The modelling approach based on financial variables to predict inflation goes back to the quantity theory of money of the early nineteenth century.[3] However, the empirical support for such approaches is still mixed and model dependent, as pointed out by Stock and Watson (2003) and Ang et al. (2007).[4] One of the reasons for this puzzle could lie in the different frequencies at which inflation and financial variables are sampled. Specifically, even if daily financial data contain useful information about current and future movements of inflation,

[3] The intuitions that the money stock anticipates inflation is much older but the first model is in Fisher (1911).

[4] Using a factor model Forni et al. (2003) present some evidence for the euro area, Giannone et al. (2008) for the US.

the fact that inflation is sampled at a lower frequency makes it hard to exploit that predictive power. Solutions to this problem consist either in sampling daily data at a lower frequency choosing, for instance, the last observation of the month, or in prefiltering daily data to convert it to a monthly frequency. Both these approaches may discard useful information and corrupt the potential relation between the variables. As we show below, it is possible to overcome this problem with a mixed-frequency-data model. This approach allows us to combine monthly determinants of inflation with the daily information coming from financial markets in the same model. In this way, we can capture the medium to low frequencies of inflation using monthly regressors without discarding any useful information coming from financial data to predict short-term movements in inflation.

3 A TWO-STEP APPROACH TO MODEL INFLATION

As key aspect of our approach is to model the low and high frequency variability of inflation separately. In this section we describe the two components that, in our view, should be used for this purpose: a core inflation index to model the medium-to-low frequencies of inflation, and daily financial variables to model the high-frequencies.

3.1 Modelling long-medium term component of inflation

In the last few years large-scale factor models have become increasingly important in the construction of reliable coincident and leading economic indicators.[5] Factor models allow us to represent parsimoniously the information embodied in a large data set by assuming that there are a few common factors that drive the dynamics of data. In particular, in the generalized dynamic factor model (GDFM) these factors are chosen in order to explain most of the variability of the data at medium and low frequency. In this way, they disentangle the medium to long-run cyclical component from the short-term dynamics and minimize the effects of idiosyncratic and transient shocks.

[5] Factor models have been used to predict macroeconomic variables in the US (e.g. Stock and Watson (2002)) and in the euro area (e.g. Altissimo et al. (2010)) and to produce measures of core inflation for the euro area (Cristadoro et al. (2005)).

More specifically, if $y_{i,t}$, $i = 1, ...N$, is a panel of interest variables, the dynamic factor model assumes that they admit the following representation

$$y_{i,t} = \chi_{i,t} + \xi_{i,t} = \sum_{j=1}^{p} b_{i,j}(L) f_{j,t} + \xi_{i,t} \tag{1}$$

where $\chi_{i,t}$and $\xi_{i,t}$ are respectively the common component and the idiosyncratic component of $y_{i,t}$. They are, by construction, unobserved, stationary and mutually orthogonal. The common component $\chi_{i,t}$ is driven by p common factors $f_{j,t}$ which are possibly loaded with different coefficients and lags. An important feature of the GDFM is that under the assumption that the variables $y_{i,t}$ are stationary, the common component can be represented as the integral of waves of different frequency, the so called spectral representation. By aggregating waves of different frequency we can decompose $\chi_{i,t}$ into the sum of a cyclical medium and long-run component $\chi^{L}_{i,t}$ and a non-cyclical, short-run component, $\chi^{S}_{i,t}$, i.e.

$$\chi_{i,t} = \chi^{L}_{i,t} + \chi^{S}_{i,t}$$

By construction $\chi^{L}_{i,t}$ is the unobserved component that drives persistent movements in $y_{i,t}$.[6] A detailed description on the estimation of $\chi^{L}_{i,t}$ can be found in Forni et al. (2002). This approach was recently employed by Cristadoro et al. (2008) to estimate a core inflation index for the euro area using a large number of both national and sectorial prices together with monetary aggregates and other macroeconomic variables.[7] In this paper we use the same core indicator since it has important advantages. First, it is constructed using data from a large panel of time series that includes sectorial and national prices as well as monetary and real variables. Then, although it is estimated using the monthly growth of consumer price index, it is as accurate as the year-on-year inflation rate in capturing medium-long term components of inflation but, unlike the latter, it is not a lagging indicator of monthly inflation. In other words, there is no trade-off between the smoothness of the indicator and the timing of the information provided, in particular in the detection of turning points. Finally, by construction it is

[6] In particular, for the core inflation index we focus on periodicity of at least one year.
[7] A complete list of the variable used in the construction of the euro area inflation index can be found in Cristadoro et al. (2005).

cleaned of measurement and idiosyncratic errors caused by sector or country specific dynamics. The estimation of the GDFM requires two conditions: all the variables of the dataset should be transformed into the same frequency; the dataset should have a balanced end of sample, i.e. no missing values in the most recent period.

In a real real-time forecasting framework this latter requirement is never met because of the different publication lags of many economic data, leading to the so called "ragged-edge data" problem. The presence of missing observation at the end of the sample, not only has important implication for the model estimation but it can also worsen the forecasting properties of the model especially in the very short-run. Several approaches have been proposed to tackle the issue of ragged-edge data. A good survey with an application to German data can be found in Marcellino and Schumacher (2010). The approach we propose is to model the medium-run component of inflation with a lagged core inflation index (from a balanced dataset) and the short-term component using high-frequency financial series that are available in real time and unaffected by revisions.

3.2 A mixed-frequency model for real-time forecasts of inflation

In this section we present the mixed-frequency model we use to forecast inflation in real time. As mentioned above, we follow the mixed data sampling regression (MIDAS) approach proposed by Ghysels et al. (2004, 2006) since it has the feature that it allows us to construct a regression model that combines both monthly and daily variables. Specifically, if $x^d_{j,t}$ is a daily variable, $\mathbf{z}_t$ a vector of monthly variables (that does not contain the core inflation index), and $B\left(L^{\frac{1}{d}};\boldsymbol{\theta}_j\right)$ a daily lag polynomial, the mixed-frequency-data model we specify for inflation π_t is given by

$$\begin{aligned}\pi_t &= \alpha_0 + \rho\pi_{t-1} + \sum_{j=1}^{N}\beta_j B\left(L^{\frac{1}{d}};\boldsymbol{\theta}_j\right)x^d_{j,t} + \alpha\chi^L_{t-h} + \gamma^{'}\mathbf{z}_t + \varepsilon_t \qquad (2)\\ &= \alpha_0 + \rho\pi_{t-1} + \sum_{j=1}^{N}\beta_j \sum_{k=1}^{D} b\left(k;\boldsymbol{\theta}_j\right)L^{\frac{k}{d}}x^d_{j,t} + \alpha\chi^L_{t-h} + \gamma^{'}\mathbf{z}_t + \varepsilon_t\end{aligned}$$

where the daily lag coefficient $b(k;\boldsymbol{\theta}_j)$ are function of a small set of parameters $\boldsymbol{\theta}_j$, N is the number of daily variables and χ^L_{t-h} is the core inflation index, described in the previous section, lagged h periods. $L^{\frac{k}{d}}$ is the daily lag operator, which ranges from the last day of month t up to D days before. This means that $L^{\frac{1}{d}}x^d_t = x^d_{t-\frac{1}{d}}$ is the last observation of month t, $L^{\frac{2}{d}}x^d_t = x^d_{t-\frac{2}{d}}$ is the next to last observation and so on. In order to avoid spurious seasonality in the model, the loading β_j should be restricted to be

$$\beta_j = (1-\rho L)\,\delta_j$$

In fact, as shown in Ghysels et al. (2004), when an autoregressive components is included in the model, π_t can be rewritten as

$$\pi_t = (1-\rho L)^{-1}\sum_{j=1}^{N}\beta_j\sum_{k=1}^{K} b(k;\boldsymbol{\theta}_j)\,L^{\frac{k}{d}}x^d_{j,t} + (1-\rho L)^{-1}\left(a_{0+}\alpha\chi^L_{t-h}+\gamma'\mathbf{z}_t+\varepsilon_t\right) \tag{3}$$

The polynomial on the daily variables $x^d_{j,t}$ is now the product of the daily lag polynomial in $L^{\frac{k}{d}}$ and the monthly lag polynomial in L, specifically $\sum_{j=0}^{\infty}(\rho L)^j$. This generates a seasonality in the response of inflation to the daily regressors irrespective of whether the daily variables have a seasonal component.

The lag coefficient $b(k;\boldsymbol{\theta}_j)$ is assumed to follow a distributed lag function that depends on a small set of parameters $\boldsymbol{\theta}_j$, i.e.

$$b(k;\boldsymbol{\theta}_j) = \frac{f(k;\boldsymbol{\theta}_j)}{\sum_{k=1}^{K} f(k;\boldsymbol{\theta}_j)}$$

This allows us to reduce the number of parameters to be estimated in the model. In general, there are many ways of parameterizing $f(k;\boldsymbol{\theta}_j)$. Two common approaches are the exponential Almon polynomial, introduced by Almon (1965), and the Beta polynomials recently proposed by Ghysels et al. (2004, 2006). The first assumes that $f(k;b_j)$ has the following functional form:

$$f(k;\boldsymbol{\theta}_j) = e^{\theta_{j,1}k+\theta_{j,2}k^2}$$

The exponential Almon polynomials can generate different shapes including increasing, decreasing, single and multiple humped patterns. The second method of parameterization is based on the Beta density function defined as

$$f(x; b_1, b_2) = B(b_1, b_2)^{-1} x^{b_1-1} (1-x)^{b_2-1}$$

where $B(b_1, b_2) \equiv \Gamma(b_1)\Gamma(b_2)/\Gamma(b_1+b_2)$ and $\Gamma(\cdot)$ is the gamma function. The Beta density is a very flexible distribution that allows many shaped weighting functions including uniform, humped and sharply decreasing (increasing) patterns.

The model can be estimated using non-linear quasi maximum likelihood. In order to improve accuracy, we assume that the loadings of the daily variables and the parameters in the vector $\boldsymbol{\theta}$ are uncorrelated and use a recursive two-stage estimation approach. We use the model to produce daily forecasts of inflation in the following way. First, we estimate the model on a daily basis and then we use the estimated parameters to make predictions on current and one-month-ahead inflation. More specifically, let's assume that we stand at the 15^{th} day of the month t, say $t(15)$. The information set available on that day consists of past values of the inflation rate[8] and the other monthly variables up to the previous month $t-1$, and all the daily variables up to the 14^{th} day of the current month, $t(14)$. If we assume, for simplicity, that we have only one daily variable, then the model we would estimate is

$$\begin{aligned} \pi_{t-1} &= \alpha_0 + \rho\pi_{t-2} + b_1 x^d_{t-1(14)} + b_2 x^d_{t-1(13)} + b_3 x^d_{t-1(12)} && (4) \\ &\quad + ... + b_{(K)} x^d_{t-1(14-d)} + \alpha\chi^L_{t-1} + \gamma z_{t-1} + \varepsilon_{t-1} && (5) \end{aligned}$$

where $x^d_{t-1(14)}$ is the daily variable available on the 14th day of the previous month $t-1$. Notice that if d is greater the 14, then we are including in the model $D-14$ daily observations of the previous month. Once the parameters have been estimated, predictions for the current inflation $\tilde{\pi}_t$ are given by

$$\begin{aligned} \tilde{\pi}_t &= \hat{\alpha}_0 + \hat{\rho}\pi_{t-1} + \hat{b}_1 x^d_{t(14)} + \hat{b}_2 x^d_{t(13)} + \hat{b}_3 x^d_{t(12)} \\ &\quad + ... + \hat{b}_K x^d_{t(14-D)} + \hat{\alpha}_t \chi^L_{t-1} + \hat{\gamma} z_{t-1} && (6) \end{aligned}$$

while one-step-ahead predictions $\tilde{\pi}_{t+1}$ are obtained by replacing π_{t-1} with

[8] The flash estimate of the HICP for the euro area is usually released on the last business day of the current month.

the current prediction $\tilde{\pi}_t$ and z_{t-1} with z_t if available. More specifically, if we stand at the 15th of September and we want to forecast current inflation $\pi_{sept.}$, we would first estimate the model using monthly data until August and daily data until the 14th of August, i.e.

$$\begin{aligned} \pi_{aug} &= \alpha z_{Jul} + \beta_1 x^d_{14aug} + \beta_2 x^d_{13aug} + \beta_3 x^d_{12aug} + ... \\ &+ \alpha_t \chi^L_{t-1} + \gamma z_{t-1} + \varepsilon_t \end{aligned} \tag{7}$$

and then forecast September inflation using the estimated coefficient and all the information available until the 15th of September, e.g.

$$\begin{aligned} \hat{\pi}_{sept} &= \hat{\alpha} z_{aug} + \hat{\beta}_1 x^d_{14sept} + \hat{\beta}_2 x^d_{13sept} + \hat{\beta}_3 x^d_{12sept} + ... \\ &+ \hat{\alpha}_t \chi^L_{t-1} + \hat{\gamma} z_{t-1} \end{aligned} \tag{8}$$

The following day we would update our information set by including the daily data of the previous day, re-estimate the model and obtain a new prediction for September inflation. One-step-ahead predictions are obtained in the same way by a direct approach. Specifically, we first estimate

$$\begin{aligned} \pi_{t-1} &= \alpha_0 + \rho \pi_{t-3} + b_1 x^d_{t-1(14)} + b_2 x^d_{t-1(13)} + b_3 x^d_{t-1(12)} \\ &+ ... + b_{(K)} x^d_{t-1(14-K)} + \alpha \chi^L_{t-1} + \gamma z_{t-1} + \varepsilon_{t-1} \end{aligned}$$

and the use the estimated parameters to make predictions for π_{t+1}

$$\begin{aligned} \tilde{\pi}_{t+1} &= \hat{\alpha}_0 + \hat{\rho} \pi_t + \hat{b}_1 x^d_{t(14)} + \hat{b}_2 x^d_{t(13)} + \hat{b}_3 x^d_{t(12)} \\ &+ ... + \hat{b}_K x^d_{t(14-K)} + \hat{\alpha}_t \chi^L_{t-1} + \hat{\gamma} z_{t-1} \end{aligned}$$

4 TWO FORECASTING APPLICATIONS IN REAL-TIME

In this section we describe our three mixed-frequency models for the euro area HICP and assess their forecasting performance in two different exercises. In the first we evaluate the monthly forecast accuracy of the three models with respect to other standard models. Then, we compare the models' daily predictions with the forecasts implied by the quotes of euro area HICP future contracts.

Each model is constructed to capture different sources of shocks that, according to economic theory, should be related to movements in inflation. All the models are characterized by the same monthly variables, chosen to model long-run developments in inflation, but different daily variables. The monthly variables are lagged inflation, the lagged yearly change of oil price and the fifth lag of the core inflation index defined in Section 3.1.[9] We included the core index lagged five because we found it has its strongest predictive power at that horizon (this is also consistent with the results of Cristadoro et al. (2005)). At the same time, this choice has two important implications. First, it significantly reduces, in the real time exercise presented below, the distortion implied by the past data revision. Second, it eliminates the ragged-edge data issue mentioned in the introduction, which can have important implications in a nowcast exercise. Finally, the inclusion of the yearly rate of change in the oil price is aimed to capture possible second round effects of energy prices in consumer inflation. The daily variables include short and long term interest rates, interest rate spreads, commodity prices and exchange rates.[10] The choice of these variables lies on their ability to capture changes in the market expectations (in particular interest rates and interest rate spreads) as well as movements in the most volatile components of inflation, specifically energy and food, which are highly correlated with the short-run changes of the overall index. Furthermore, market expectations might also anticipate changes in indirect taxes or other government measures that cannot be easily accounted for in a pure backward-looking model.

More specifically, in the first model (M1) we include the short-term rate and changes in interest rate spread and in oil future prices. These variables should reflect both changes in expectations and future expected movements in energy prices. The second model (M2) is designed to capture recent shocks coming from outside the euro area that are not yet embodied in lagged core inflation. For this reason we considered changes in the wheat price, in the oil futures quotes and in the exchange rate. Finally, in the third model (M3)

[9] The monthly inflation is defined as year on year change of the euro area HICP index.

[10] The short-term rate is the 3-month Euro LIBOR rate while the long-term rate is the yield on the 10 year Bund. The interest rate spread is constructed as the difference between the 10-year German Bund and 3-month German interbank rate. The quotes of the future contracts and the spot oil prices refer to the Brent oil. The exchange rate is the effective rate (EER22). Daily data for the financial variables are taken from DATASTREAM; euro area HICP CME future quotes are extracted from Bloomberg, and monthly data for the HICP and interest rates from the ECB Statistical Data Warehouse (SDW).

we only focus on inflation expectations embodied in the interest rates and for this reason we include long-term rates and changes in the interest rate spreads and in the short term rate. The choice of three different models with a small number of daily variables is based on the need to keep the model parsimonious and, at the same time, to identify different sources that could partially explain short-run developments in euro area HICP.[11]

4.1 Real-time forecasts of monthly inflation

In this section we present two empirical exercises to assess in pseudo real time the forecasting ability of our models compared with monthly VAR models and market daily expectations. Even if we did not take into account the issue of data revision, the difference from a pure real-time application is negligible for two reasons. First, the use of the fifth lag of the monthly core inflation implies that we can estimate it with a balanced dataset and that the effects of data revision should be negligible. Second, the daily financial variables are available in real time and not subject to revisions.

In the first exercise we compute the root mean squared forecast error (RMSFE) of our mixed-frequency models and compare them with those from univariate and multivariate models. In particular, among the univariate models we consider a random walk, an AR model and an ARMA model whose order is chosen with the Schwarz criterion. In order to assess the contribution of the daily variables to improving the forecast accuracy of our three M-models, we consider two VARs: the first has the same monthly variables as our models (inflation, core inflation and the oil price) therefore we call VAR_MIDAS; the second also includes the short-term interest rate, because of their economic relevance, therefore we call VAR_ECON.

Since these models produce monthly forecasts, the comparison has been done in the following way. For each day of each month we estimate our three M-models, make predictions for current and one-month-ahead inflation and finally take the monthly average of the daily forecasts. Our data set starts in May 1992 therefore we used ten years of data as a burning period and then we run recursive forecasts for monthly inflation starting from May 2002 until September 2007. The RMSFE for all models are showed in Table 1. All the

[11] In the early stage of our work we noticed that the inclusion of a large number of daily variables in the same model could lead to highly volatile parameter estimates. A possible solution that could be employed in a future extension of this work would be the construction of daily financial factors in line with Andreou et al. (2010).

mixed-frequency data models outperform the univariate models as well as the VAR models. The average reduction in the forecast errors is more than 20% both for current and one-month-ahead inflation. Given that VARs use the same variables as our model, the better forecasting performance of our models implies that the daily variables have significant predictive content. Although this result can be partially explained by the larger number of variables in the mixed frequency models, it must be noticed that our models impose much stronger restrictions on the data dynamics compared with the unrestricted VARs, which allows for important feedbacks among the variables. Finally, among the mixed-frequency models, M2 performs the best, showing that, in the sample period we considered, movements in commodities prices and exchange rates seem to be good predictors of the short-term developments in the EA HICP. This is not surprising if we consider the euro sharp depreciation in the first few years after its introduction.

4.2 Model forecasts vs market expectations

In the second application, we fully exploit the high-frequency structure of our models and run a daily forecasting exercise. Given the lack in the literature of benchmark models for daily forecasts of inflation, we decided to compare predictions of our models directly with market expectations, in particular with the inflation expectations extracted from the economic derivatives. Economic derivatives are securities whose payoff is dependent upon macroeconomic data releases. They were recently introduced and have become popular for their ability to mitigate some of the market risks found in standard instruments linked to inflation, such as the US TIPS bonds. Since their yields are tied to future data releases of a certain macroeconomic variable, they can be also considered as a good measure of the market beliefs about future realizations of the economy. Wolfers and Gurkaynak (2006) shows in fact that market expectations extracted from economic derivatives are more accurate than survey data in predicting many US variables.

In this exercise we focus on the future contracts on the euro area inflation, defined as the yearly percentage change of the harmonized index of consumer price (HICP) excluding tobacco, released monthly by Eurostat around 15 days after the end of the month.[12] These future contracts were

[12]The impact of tobacco on the overall index is negligible and estimated to be at most 2.5%. The correlation between the two indexes is 0.94.

introduced in 2005 and are traded daily on the Chicago Mercantile Exchange (CME).[13] Since we are using contracts that are very close to maturity date, we should expect liquidity and risk premia to be negligible especially at such short term horizon (see Piazzesi and Swanson 2008). Figure 1 shows the expected inflation rates extracted from the futures contracts versus the realized inflation rate: they track quite accurately the annual growth rate of the euro area HICP.

Another popular way in the literature to extract market expectations of inflation is to look at the so called break-even inflation rate. This is defined as the difference between yields of inflation-linked bonds and those of fixed rate bonds. For our empirical analysis we preferred to use the future contracts rather than the break-even inflation rate since they seem to be more accurate at predicting inflation at such short-term horizons. Furthermore, inflation-indexed bonds are influenced by microstructure and tax factors that are hard to quantify.

The empirical exercise is conducted as follows. For each day of the sample we estimate the model and generate predictions for current and one-month-ahead inflation.[14] In particular, our aim is to forecast the inflation rate that will be released at the end of the current month and of the following one. In Figure 2 we show the boxplot for the daily forecast errors of the futures contracts and for those of the three mixed-frequency models. The left-hand panel refers to the distribution of the forecast errors for the current month, the right-hand panel for one month ahead. All the mixed-frequency models produce not only more accurate but also less volatile forecasts, especially for current inflation. The improvement in the forecast accuracy is on average around 30% for current inflation and somewhat smaller for one-month-ahead inflation, as shown in Table 2, where we report the RMSFE for the three models and the predictions of the futures contracts. In contrast with what we found in the previous exercise, M1 seems to perform slightly better than the other models. Since the forecast periods of the two exercises overlap only in the last two years of the sample this could be interpreted as indicating

[13] A detailed description of the contract can be found Grannan and Srinivasan (2007). To date there are more than 1.5 million contracts for the euro area HICP traded every day. The bid-ask spread for these contracts is reasonably tight, implying a low liquidity premium.

[14] As in the previous application, the estimation period starts in May 1992, while the forecasting exercise begins in October 3 of 2005, because of data constrains in our economics derivatives data base.

an increased importance of interest rates in explaining inflation dynamics compared with the exchange rate. This is not surprising if we consider the relative stability of the exchange rate during this period compared with the significant depreciations of the euro that occurred between 2002 and 2005.

In Figure 1 we plot the evolution of the model daily forecasts. The graph shows the daily forecasts of model one (M1), the inflation rates extracted from the quotes of the futures contracts and the actual inflation rate (our target). Apart from a couple of episodes when our model could not capture a sudden drop in inflation, it seems they track inflation more closely than market expectations. We computed the Diebold-Mariano (DM) test to assess whether this improvement in the forecasts is statistically significant. The DM statistics in Table 3, adjusted as in West (1996), show that model predictions for current inflation are significantly more accurate than those of the derivatives; however, there is no strong evidence for one month ahead. In order to assess the goodness of this result, we also computed the Chong and Hendry (1986) forecast encompassing regressions to check whether our models encompass the information of the derivatives over future releases of inflation.[15] Specifically, we construct a daily measure of inflation π_t^d by keeping monthly inflation π_t constant for each day of the month t and regress it over a constant, the daily future contracts expectations D_t^d and our daily model projections $\hat{M}_t^d$:

$$\pi_t^d = \beta_0 + \beta_1 D_t^d + \beta_2 \widehat{M}_t^d \quad (9)$$

The results in Table 4 show a very different picture for the two forecasting horizons. For current inflation, the coefficient β_2 is much bigger than β_1 confirming that the model forecasts are more accurate than future contracts do. However, both coefficients are significant, suggesting that future contracts still have a small but significant predictive content that is not captured by our models. This implies that we could further reduce the forecast error by combining the predictions from our models with those from the future contracts. The best approach to extract their forecast ability would be to use them directly in our mixed frequency models. However, since they have been quoted only since 2005, their sample length is too short for model estimation. Therefore, we preferred to resort to a forecast combination approach.

[15] As recently shown by Busetti et al. (2009), the Diebold and Mariano (1995) test can be characterized by low size compared to forecast encompassing tests.

In Table 5 we show the RMSFE for the combined predictions obtained with estimated weights and equal weights. Unsurprisingly, we find that combining the two forecasts brings a non negligible reduction in the RMSFE especially for the one-month-ahead predictions.

5 CONCLUDING REMARKS

In this paper we present a simple model to forecast euro area inflation in real time. It is based on a mixed-frequency model that combines two important components to forecast inflation at short-term horizons: a monthly core inflation index derived from a dynamic factor model that captures persistent changes in inflation, and daily financial variables that are used to extract timely information on the most recent shocks. Financial variables are useful in our context as they are forward looking, easy to observe in continuous time and not subject to revisions, unlike most of the hard data. We use the MIDAS approach, proposed by Ghysels et al. (2004, 2006), to keep the mixed-frequency model tractable and parsimonious. Our approach presents two main advantages. First, the mixed frequency structure of our model allows us to produce real-time forecasts of inflation that are updated with the latest market information. Second, it significantly reduces the unbalanced (ragged-edge) dataset problem typical of the dynamic factor models, which can produce considerable bias in short-term forecasts. Given the lack of daily models for inflation, we assess the forecasting ability of our models with respect to multivariate reduced form monthly models and daily economic derivatives on euro area inflation. We find that the inclusion of the daily variables helps to reduce the forecasting errors relative to models with only monthly variables and forecasts extracted from daily economic derivatives. We find that the combining in a mixed frequency model both daily finacial variables and monthly economic indicators helps to reduce the forecasting errors relative to models with only monthly variables and forecasts extracted from daily economic derivatives. The predictive ability of daily asset and commodity prices for current inflation rate is consistent with forward-looking behavior in the price setting mechanism of firms that could be interesting to study with a structural economic model.

REFERENCES

Almon, S. (1965), "The distributed lag between capital appropriations and expenditures," Econometrica, 33, 178-196.

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi and G. F. Veronese, (2010), "New Eurocoin: Tracking Economic Growth in Real Time," The Review of Economics and Statistics, 92(4) .

Andersen, T. G., T. Bollerslev, F. X. Diebold and C. Vega, (2003)," Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," American Economic Review, 93(1).

Andreou, E., E. Ghysels and A. Kourtellos (2010), "Should macroeconomic forecasters use daily financial data and how?," paper presented at 6th Workshop on Forecasting Techniques Forecasting Real time and Survey data, Bundesbank, EABCN and ECB, Frankfurt, 5-6 March.

Ang, A., G. Bekaert and M. Wei, (2007), "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?," Journal of Monetary Economics, No.54.

Aruoba, S. B., F. X. Diebold and C. Scotti, (2009), "Real-time Measurement of Business Conditions," Journal of Business and Economic Statistics, No. 27.

Bernanke, B.S. and K. N. Kuttner, (2005), "What Explains the Stock Market's Reaction to Federal Reserve Policy?," The Journal of Finance No. 60.

Busetti, F., J. Marcucci and G. Veronese, (2009), "Comparing Forecast Accuracy: a Monte Carlo Investigation," Temi di discussione (Economic working papers) No. 723, Banca d'Italia.

Chong, Y.Y. and D.F. Hendry, (1986), "Econometric Evaluation of Linear Macro-Economic Models," Review of Economic Studies, No. 53.

Clements, M. P., and A. B. Galvão, (2008), "Macroeconomic Forecasting with Mixed Frequency Data: Forecasting Output Growth in the united States", Journal of Business and Economic Statistics, 26, No.4

Cristadoro, R., M. Forni, L. Reichlin and G. Veronese, (2005), "A Core Inflation Indicator for the Euro Area," Journal of Money, Credit and Banking, No. 37(3).

Cristadoro, R., G. Saporito, and F. Venditti, (2008), "Forecasting inflation and tracking monetary policy in the euro area - does national information help?" ECB Working Paper, No. 900.

Dielbold, F. X. and R.S. Mariano, (1995), "Comparing predictive accuracy", Journal of Business and Economic Statistics, Vol. 55.

Fisher, I., (1911), The Purchasing Power of Money, Macmillan, New York.

Forni, M., M. Hallin, M. Lippi and L. Reichlin, (2002), "The Generalized Factor Model: Identification and Estimation," The Review of Economics and Statistics, No. 82(4).

Forni, M., M. Hallin, M. Lippi and L. Reichlin, (2003), "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?," Journal of Monetary Economics, No. 50(6).

Ghysels, E., P. Santa-Clara and R. Valkanov (2004), "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, (2009), "Forecasting Professional Forecasters," Journal of Business, Economics and Statistics, No. 27.

Giannone, D., L. Reichlin and D. H. Small, (2008), "Nowcasting GDP and Inflation: the Real-time Informational Content of Macroeconomic Data Releases," Journal of Monetary Economics, No. 55.

Grannan, L., and S. Srinivasan, (2007), "Exchange-Traded Inflation Derivatives: A New Asset Class," Chicago Mercantile Exchange.

Marcellino, M., and C. Schumacher, (2010), "Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP," Oxford Bulletin of Economics and Statistics, 72(4).

Mittnik, S. and P. Zadrozny, (2004), "Forecasting Quarterly German GP at Monthly Interval Using Monthly IFO Business Condition Data," CESifo Working Paper, No. 1203.

Piazzesi, M., and E. T. Swanson, (2008), "Futures Prices as Risk-adjusted Forecasts of Monetary Policy," Journal of Monetary Economics, No. 55.

Stock, J., and M. Watson (2002), "Macroeconomic Forecasting Using Diffusion Indexes," Journal of Business & Economic Statistics, American Statistical Association, No. 20(2).

Stock, J.H., and M.W. Watson, (2003), "Forecasting Output and Inflation: The Role of Asset Prices," Journal of Economic Literature, No. 41.

Stock, J., and M. Watson, (2006), "Macroeconomic Forecasting Using Many Predictors," in Handbook of Economic Forecasting, G. Elliott, C. Granger and A. Timmermann (eds.), North Holland.

West, K. (1996) "Asymptotic Inference About Predictive Ability," Econometrica No.64.

Wolfers, J., and R. S. Gurkaynak, (2006), "Macroeconomic Derivatives: An Initial Analysis of Market-Based Macro Forecasts, Uncertainty and Risk", CEPR Discussion Paper No. 5466.

Woodford, M., (2003), Interest and Prices:Foundations of a Theory of Monetary Policy, Princeton University Press.

APPENDIX

Table 1 **Monthly forecasting accuracy**

(RMSFE of recursive forecasts from 2002:5 to 2007:9)

	steps ahead	
	0	1
Compound models		
M1	0.158	0.226
M2	**0.148**	**0.208**
M3	0.163	0.216
Univariate models		
RW	0.185	0.261
AR(1)	0.181	0.250
ARMA(2,1)	0.184	0.248
Multivariate models		
VAR_MIDAS	**0.169**	**0.234**
VAR_ECON	0.173	0.244

Legend:

RW = random walk of inflation

VAR_MIDAS= VAR(2) with inflation, core inflation and oil prices

VAR_ECON= VAR(3) with inflation, core inflation, oil prices and interest rates.

Notes: in each entry Root Mean Square Forecast Error (RMSFE) is reported. Entries in bold indicate the lowest RMSFE. The first estimation period starts in may 1992 and ends in may 2002. Subsequent estimates follow the recursive scheme (keeping fixed the starting date). Daily forecasts for M1, M2 and M3 are aggregated to produce monthly predictions.

Table 2 **RMSFE of daily predictions**

(RMSFE of recursive forecasts from 2002:5 to 2007:9)

	Econ deriv	M1	M2	M3
current month	0.166	0.123	0.132	0.140
one month ahead	0.233	0.219	0.224	0.211

Notes: Each entry reports the RMSFE of the different predictors. The model's prediction errors refer to the recursive unconditional out of sample forecast. Sample period: October 3, 2005 - September 30, 2007.

Table 3 **Comparing predictive accuracy: Diebold Mariano forecasting test**

	current month			one month ahead		
	M1	M2	M3	M1	M2	M3
MSE	0.012 **(2.883)**	0.010 **(2.277)**	0.008 (1.618)	0.006 (0.763)	0.004 (0.434)	0.010 (0.943)
MAE	0.044 **(3.754)**	0.037 **(3.061)**	0.035 **(2.786)**	0.015 (0.889)	0.028 (1.748)	0.032 (1.909)

Notes: the test is on the difference between derivative and model prediction errors. The null hypothesis is that the expected error of the competing forecast is equal to that of the futures market. In brackets the heteroskedasticity and autocorrelation consistent (Newey West) t-stat of the coefficients. Bold t-statistics imply significance at the 5 % level. Sample period: October 3, 2005 - September 30, 2007.

MSE: test on the squared residuals

MAE: test on the absolute residuals

Table 4 **Forecast encompassing test: combination weights**

	current month			one month ahead		
	M1	M2	M3	M1	M2	M3
Coefficients:						
constant	-0.136	-0.188	-0.228	0.044	0.176	0.036
	(-1.366)	(-2.113)	(-2.262)	(0.296)	(1.277)	(0.241)
derivatives	0.297	0.351	0.414	0.389	0.388	0.384
	(3.083)	(3.670)	(4.017)	(4.835)	(4.650)	(4.590)
model	0.769	0.634	0.691	0.609	0.517	0.588
	(9.749)	(8.893)	(7.821)	(6.356)	(6.442)	(5.997)
Test a diagnostics:						
Wald F statistic	38.23	30.78	23.76	22.27	21.97	21.15
P value, %	0	0	0	0	0	0
Test b diagnostics:						
Wald F statistic	3.37	5.74	6.23	15.86	12.33	7.65
P value, %	1.8	0.0	0.2	0	0	0

Notes: (1) Equation (7) forecast encompassing test on the daily forecasts
(2) Test a: the null hypotesis is that future forecasts encompass model forecasts
(3) Test b: the null hypotesis is that model forecasts encompass future forecasts
(4) In brackets: heteroskedasticity and autocorrelation consistent (Newey West) t-stat of the coefficients.

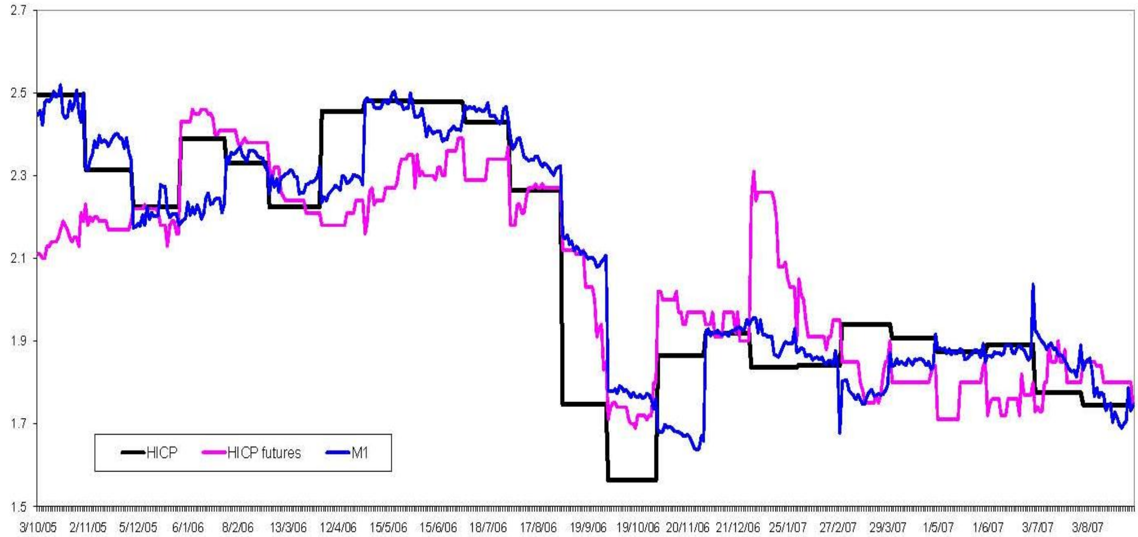
Table 5 RMSFE of combined daily predictions

	Econ deriv	Combined predictions		
		M1	M2	M3
Equal weights				
current month	0.166	0.124	0.127	0.130
one month ahead	0.233	0.192	0.195	0.190
Estimated weights				
current month	0.166	**0.117**	0.123	0.128
one month ahead	0.233	0.191	0.195	**0.189**

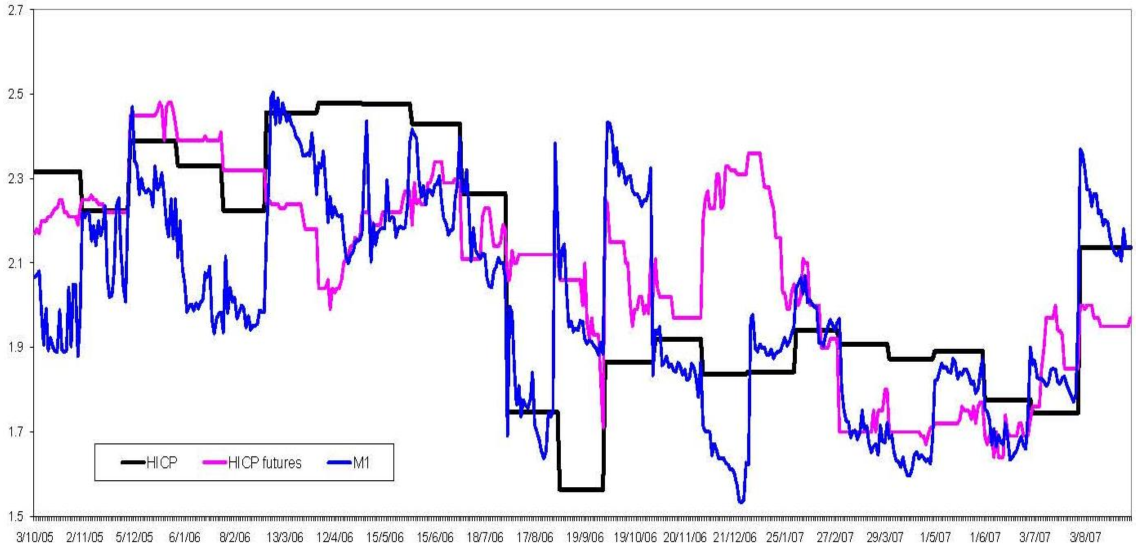
Notes: Estimated weights are those of forecast encompassing regressions (Table 4).

Figure 1 **Daily forecasts**

current month

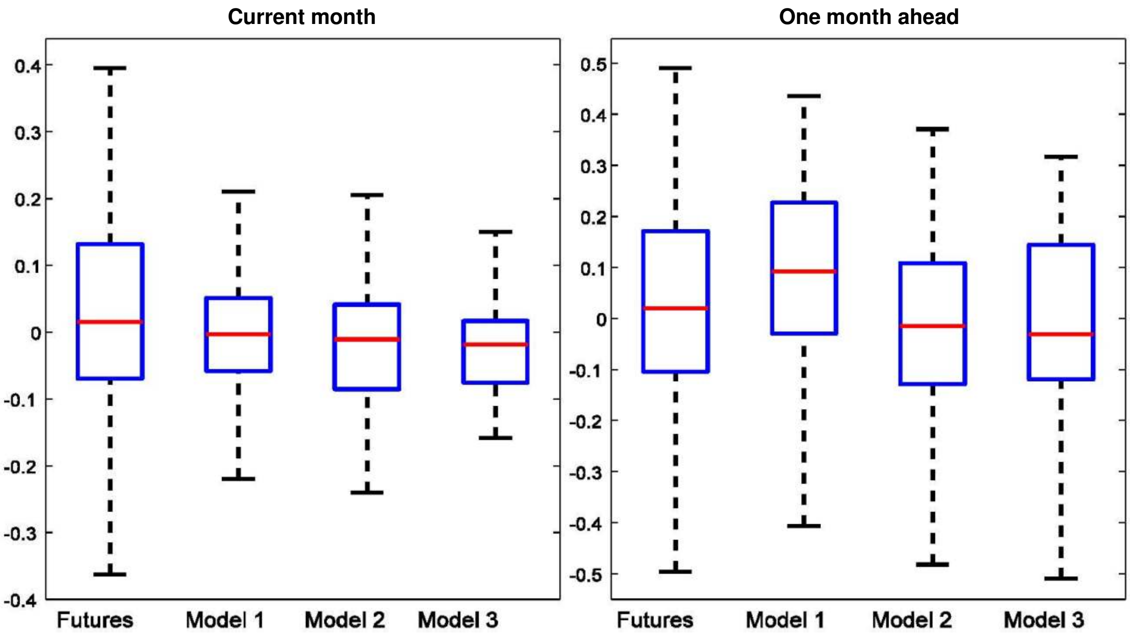


one month ahead



Legend:
HICP futures= Monthly inflation rate implied in the daily HICP future contracts (source: Bloomberg).
HICP = HICP inflation rate projected on daily data (source: Eurostat).
m1 = Daily inflation predictions of model 1.

Figure 2 **Box-Plots of daily forecast errors**



Legend:
Futures = Forecast errors of the daily HICP future contracts (source: Bloomberg)
Model 1,2,3 = Forecast errors of the MIDAS models.
The blu box portion represents the first and third quartiles. The median is the red line through the center of the box. The staple is a black line drawn at the last data point within (or equal to) each of the inner fences.
Sample period: October 3, 2005 - September 30, 2007.

Ministry of Economy and Finance

Department of the Treasury

Directorate I: Economic and Financial Analysis

Address:
Via XX Settembre, 97
00187 - Rome

Websites:
www.mef.gov.it
www.dt.tesoro.it

e-mail:
dt.segreteria.direzione1@tesoro.it

Telephone:
+39 06 47614202
+39 06 47614197

Fax:
+39 06 47821886