International Journal of Forecasting 29 (2013) 664–675

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Now-casting inflation using high frequency data

Michele Modugno

Université libre de Bruxelles, Belgium

ARTICLE INFO

Keywords:
Factor models
Forecasting
Inflation
Mixed frequencies

ABSTRACT

This paper proposes a methodology for now-casting and forecasting inflation using data with a sampling frequency which is higher than monthly. The data are modeled as a trading day frequency factor model, with missing observations in a state space representation. For the estimation we adopt the methodology proposed by Bańbura and Modugno (2010). In contrast to other existing approaches, the methodology used in this paper has the advantage of modeling all data within a single unified framework which allows one to disentangle the model-based news from each data release and subsequently to assess its impact on the forecast revision. The results show that the inclusion of high frequency data on energy and raw material prices in our data set contributes considerably to the gradual improvement of the model performance. As long as these data sources are included in our data set, the inclusion of financial variables does not make any considerable improvement to the now-casting accuracy.

1. Introduction

Forecasting inflation is of interest to both market practitioners and central banks. Market practitioners monitor macroeconomic releases continuously, among them inflation, in order to update their expectations on the future developments of macroeconomic fundamentals, and thus adjust their investment strategies. Central banks are charged with the task of guaranteeing price stability, and therefore routinely monitor inflation expectations and update inflation forecasts in their efforts to pinpoint the current inflation developments and to understand the underlying forces that may jeopardize price stability. In this context, now-casting inflation can help policy makers to make more accurate and timely monetary policy decisions.

This paper provides an econometric framework which allows interested parties to update their inflation forecasts continuously, following growing amounts of incoming information on a wide range of relevant available data series. It also allows the effects of different data groups on forecast revisions to be disentangled. The relevant data groups include variables which are highly correlated with inflation and which are released earlier than the relevant inflation releases. Moreover, they include data which are sampled at different frequencies, which allows information to enter the model in a sequential manner, as information updates become available.

The empirical analysis focuses on US CPI inflation. US CPI data are typically released at the middle of the month following the reference month. However, a wide range of data are released during the reference month at weekly or daily sampling frequencies, which carry valuable information on consumer prices. This information and the early signals that it contains can be useful for improving the accuracy of the predicted actual inflation and its future developments.

This approach was inspired by the relevant GDP now-casting literature. Indeed, although now-casting inflation is a novel idea, there is a rather long history of studies focusing on now-casting GDP. GDP is a quarterly variable released with a substantial time delay (e.g., one month after the end of the reference quarter for the US GDP), and several monthly indicators are released in the meantime. The seminal paper by Giannone, Reichlin, and Small (2008) for the US, and its first applications for the euro area, by Angelini, Bańbura, and Rünstler (2007) and Bańbura and

E-mail address: mmodugno@ulb.ac.be.

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2012.12.003

Rünstler (2007), show that using monthly indicators is crucial to accurate now-casting of the GDP.

Turning our focus to forecasting inflation, the topic has attracted the attention of various academics in the past. Attempts to forecast inflation include, among others, Cecchetti, Chu, and Steindel (2000) (who use univariate leading indicators models), Stock and Watson (1999) (who adopt factor models), Stock and Watson (2003) (who exploit asset prices in regression models) and Banerjee and Marcellino (2006) (who assess the relative forecast accuracy among an automated model selection procedure, a factor model, and single-indicator-based forecast pooling). However, these papers do not take into account the real-time availability and the mixed frequency nature of the data being used. Moreover, they focus mainly on long- rather than short-horizon predictions.

More recently, two different approaches have been proposed for the use of high frequency indicators in now-casting/forecasting inflation. In the first one, Lenza and Warmedinger (2011) use monthly, weekly and daily data for now-casting/forecasting inflation with a factor model. In order to add weekly and daily data to the monthly data, they use univariate autoregressive models to produce forecasts of the higher frequency data in order to fill in the missing data within each month. Once these forecasts have been produced, the available information and the forecasts produced are aggregated for each month, to obtain the monthly counterparts of the daily and weekly data. Then, a factor model is employed in order to produce predictions of inflation. In the second approach, a new generation of models, the Mixed Data Sampling Regression Models (MIDAS), originally proposed by Ghysels, Santa-Clara, and Valkanov (2004), have been used by Monteforte and Moretti (2010) to forecast inflation using a two-step approach. They extract principal components from a large sample of daily financial variables, and then use them in the forecasting equation for the target variable. In order to prevent over-parametrization, the MIDAS approach assumes that the response to the high frequency explanatory variables follows a distributed lag polynomial.

In contrast to the above-mentioned procedures, the methodology used in this paper models data with mixed frequencies within a single unified, coherent framework. The data are modelled as a trading day frequency factor model with missing observations, cast in a state space representation. The estimation adopts the methodology proposed by Bańbura and Modugno (2010), based on the EM algorithm for unobserved components models originally presented by Watson and Engle (1983). Doz, Giannone, and Reichlin (2006) later showed that this algorithm allows the estimation of factor models by maximum likelihood using large panels of data. Taking this into account, Bańbura and Modugno (2010) generalize Watson and Engle's (1983) methodology in order to deal with large panels which are characterized by arbitrary patterns of missing data.

In this manner, the econometric methodology proposed in this paper fully exploits the co-movement of data with different frequencies, and stresses the importance of timeliness. That is, our methodology allows the information content of each added variable to be preserved, as it avoids the need to introduce model-based forecasts for missing data and aggregation procedures which combine observed and forecasted data. In practice, this method also allows us to assess the importance of timeliness in the information that enters in the models when forecasting inflation, which is typically more pronounced in higher frequency data. This is achieved by comparing the forecast performance of a factor model which uses only monthly data with the forecast performance of the proposed factor model, which also uses weekly and daily data. In this paper, weekly data include Weekly Retail Gasoline and Diesel Prices (WRGDP), and daily data include the World Market Prices of Raw Materials (RMP) and some of the most important financial variables.

Finally, the proposed methodology also allows the unpredicted information content of the releases (news) to be linked with the forecast revisions of the target variable directly. Specifically, the proposed framework offers the possibility of forecasting all of the variables involved individually (unlike the MIDAS procedure, where the higher frequency variables are considered exogenous). Model-based news from each variable introduced can then be defined as the difference between the data released and the respective model predictions. We can thus disentangle the differing impacts of various releases on the forecast revision, which helps to increase our understanding of the variables which affect the forecast outcome. This is an additional feature that this paper aims to introduce for the now-casting of inflation.

The empirical evidence shows that higher frequency data, which are more timely than lower frequency ones, are necessary in order to produce more accurate inflation forecasts. The model suggested by this paper significantly outperforms (at short horizons, i.e., zero and one month(s) ahead) the model using lower frequencies only, as well as standard benchmark models for forecasting inflation (the Random Walk and Integrated Moving Average models).

Moreover, the empirical evidence evaluates the subcomponents of inflation as well as the importance of the different data groups in forecasting inflation. Looking at the forecasting performances of the subcomponents of inflation, our results suggest that the forecast improvement is due, to a larger extent, to the more accurate forecasts of the energy component and the transport component. Furthermore, the inclusion of RMP and WRGDP improves the accuracy of the model significantly. However, as in Stock and Watson (2003), the evidence is not equally supportive of financial data, which appear to have only a marginal contribution, when utilized together with the other variable groups.

The paper is organized as follows: Section 2 describes the data, and Section 3 introduces the model and the estimation methodology. Section 4 presents the results from the forecast exercise, after which Section 5 explains the concept of news and presents an illustrative example. Finally, Section 6 concludes.

2. Data

The data included in the analysis have two main characteristics: a high correlation with inflation, and timeliness,