Universitat Autònoma de Barcelona

Essays on Macroeconomic Forecasting

Alex Tagliabracci

Doctoral Thesis:
Essays on Macroeconomic Forecasting

Alex Tagliabracci

Advisor: Prof. Luca Gambetti

Department d'Economia i d'Història Economica
Universitat Autònoma de Barcelona

This dissertation is submitted for the degree of
Doctor of Philosophy at the International Doctorate in Economic Analsysis (IDEA)

September 2018

To whom who believed and still believe in me

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

This work represents my own view and the opinions expressed in this thesis are those of the author and do not necessarily reflect those of the Bank of Italy, the Eurosystem and any istitution I have been affiliated in the past.

Alex Tagliabracci
September 2018

Acknowledgements

These last five years have been a long but marvelous journey. I will try to name few people that made this experience possible.

First of all, I want to thank Prof. Mario Forni for his help before starting my PhD path. He not only incetivates me to keep studying after my master, but he also made me realize what my main research interests were. I am deeply grateful to have had him as a supervisor.

I am grateful to the Universitat Autònoma de Barcelona (UAB) and all its professors from the International Doctorate in Economic Analysis (IDEA). The first two years of PhD were incredibly demanding but after all extremely rewarding. I have to thank UAB and MOVE for their financial support, the dedication and the effort I could spend on my research would have been impossible without such financial support.

A big thank goes to all the colleagues and friends at UAB that shared this path with me. In particular, thank to Alessandro, Anna, Annalisa, Codruta, Davide, Josè. Thanks for keeping me going on every day. Thanks for sharing laughter and tears over the problem sets with me. A special thank to Lavinia for being the friend you need next to you not only during a PhD.

During the last three years, I had the opportunity of doing different working experiences from which I learnt a lot. I would like to thank Eva and Pep for their hospitality at Banc Sabadell, I understood the importance of that experience only afterwards.

Thanks to my former colleagues at Rokos Capital Management. Thanks to Jacques for believing in me from day one. Thanks to André, Seth and Silvio for teaching so many things that can not be found in books. Thanks to Aleksandra and Sam, best deskmates you can desire for. Thanks to Luca, our friendship began in London and I hope that the third chapter of this thesis will not be our last work together. A final thanks to Antonello D'Agostino, I benefited so much from working with you and I can not describe in words how important you were in my career. I will thank you forever.

I thank the European Central Bank (ECB) and my former colleagues for giving me the possibility to work with them. In particular, I thank Alberto Musso for being an incredible motivator during my first staying at ECB. I thank Gerrit Koester, it was a pleasure to work with you. Andrea, Laura, Marco, Marzio and all the other friends that I met at ECB made my staying in Frankfurt a wonderful experience. A special thank to Chiara Osbat, I learn so much from working with you.

I thank my collegues at Bank of Italy for accodomodating my needs and giving me the possibility of concluding this thesis, your availability has been very much appreciated.

I thank to my PhD supervisor Luca Gambetti for teaching me macroeconometrics and giving me the possibility to make the experiences described before. Thank you.

I also thank my friends in Pesaro, any time I came back home you were ready to fill my batteries with some fresh energy.

Last but not least, thanks to my family for their unconditional and constant support. I dedicate this thesis to them.

Abstract

This thesis is a collection of three empirical essays with a focus on forecasting. The first chapter focuses on an important policy task as forecasting inflation. The work aims to investigate how the dynamics of the business cycle may impact the distribution of inflation forecasts. The second chapter considers two econometric models used in the nowcasting literature and propose a comparison with an application to the Italian GDP. The last chapter is centered around forecasting the effects of macroeconomic data releases on the exchange rates. Now I will analyze more in details these works.

The first chapter studies how the business cycle affects the conditional distribution of euro area inflation forecasts. Using a quantile regression approach, I estimate the conditional distribution of inflation to show its evolution over time allowing for asymmetries across quantiles. I document the evidence of downside risks to inflation which vary in relation to developments of the state of the economy while the upside risks remain relatively stable over time. I also find that this evidence partially characterizes the corresponding distribution derived from ECB Survey of Professional Forecasters.

The second chapter proposes two multivariate econometric models that consider two important characteristics in the nowcasting literature, as timely and high frequency data, to predict Italian GDP, namely a dynamic factor model and a mixed-frequency Bayesian VAR. A pseudo out-of-sample exercise shows three main results: (i) both models considerably outperform a standard univariate benchmark; (ii) the dynamic factor model turns out to be more reliable at the end of the forecasting period while the mixed-frequency BVAR appears superior with an incomplete information set; (iii) the overall forecasting superiority of the dynamic factor model is mainly driven by its ability in capturing the severity of recession episodes.

Finally, the third chapter, jointly written with Luca Brugnolini and Antonello D'Agostino, investigates the possible predictability of macroeconomic surprises and their effects on the exchange rates. In particular, we analyze two of the most important data releases that impact the US financial market, namely the change in the level of non-farm payroll employment (NFP) and the manufacturing index published by the

Institute for Supply Management (ISM). We examine the unexpected component of these two, as measured by the deviation of the actual release from the Bloomberg Consensus. We label it as the market surprise, and we investigate whether its structure is partially predictable and in which cases. Secondly, we use high-frequency data on the eurodollar as a laboratory to study the effect of these surprises. We show in a regression framework that although the in-sample fit is sufficiently good, the performance deteriorates in an out-of-sample setting because a naive model can hardly be beaten in a sixty-minute window after the release. Finally, we demonstrate that under certain circumstances there is some structure that can be exploited and we provide a framework to take advantages of it.

This work represents my own view and the opinions expressed in this thesis are those of the author and do not necessarily reflect those of the Bank of Italy, the Eurosystem and the istitutions I have been affiliated in the past.

All remaining errors are my own.

Table of contents

List of figures

List of tables

Chapter 1

The Vulnerability of Euro Area Inflation Forecasts

Abstract

Macroeconomic conditions are generally important determinants of the inflation outlook. This paper studies how the business cycle affects the conditional distribution of euro area inflation forecasts. Using a quantile regression approach, I estimate the conditional distribution of inflation to show its evolution over time allowing for asymmetries across quantiles. I document the evidence of downside risks to inflation which vary in relation to developments of the state of the economy while the upside risks remain relatively stable over time. Interestingly, this evidence can partially be found in the corresponding conditional distribution derived from the ECB Survey of Professional Forecasters.

JEL classification: C32, E31, E32, E37

Keywords: inflation, quantile regression, conditional distribution, downside risks

“If we look at the uncertainty dimension of inflation, we see two phenomena. The first is that tail risks have disappeared. [...] And we also see that the uncertainty about the path of inflation has also decreased."

Mario Draghi, 08 June 2017

1.1 Introduction

Predicting inflation is certainly of primary importance. Over the last years the literature on forecasting inflation has showed an increased interest on the level of uncertainty attached to point estimates, which is generally captured by means of the predictive density (e.g. Elliott and Timmermann (2008)). An accurate characterization of the degree of perceived risks about future inflation is undoubtedly relevant, both for economic agents and policymakers. Indeed, professional forecasters and more in general market operators who are asked to participate in economic surveys, have to provide not only their point estimates but also the probability distribution around their forecasts. Similarly, policymakers regularly evaluate the likelihood of different scenarios, such as deflation or high-inflation rate, to gauge the risk associated to future inflation, which is a major threat to policy effectiveness.

This paper contributes to the literature on forecasting inflation by providing a comprehensive study on how macroeconomic conditions shape the conditional distribution of inflation forecasts. First, a quantile-regression approach is adopted to characterize the effects of changes in the current state of the economy with respect to the entire distribution of inflation. Then, the quantile function is used to estimate the conditional distribution of inflation forecasts. This approach follows Adrian et al. (2016) who focus on the vulnerability of US gross domestic product (GDP). Their approach fits with the purpose of this paper because it allows to investigate the properties of the conditional distribution by providing an extensive analysis of its characteristics, focusing not only on the first two moments, as done by most of the literature, but rather on the entire shape of the distribution.

Previous studies have already analyzed the distribution of inflation forecasts but mainly focusing on different features. Some examples are Tsong and Lee (2011) who show asymmetric inflation dynamics for 12 OECD countries, Tillmann and Wolters (2015) that find a structural break in persistence at all quantiles of the US inflation in the early 80s and Manzan and Zerom (2013) who challenged the view that random-walk models are more accurate for forecasting inflation by exploring the power of leading indicators of economic activity as valuable predictors, especially at the tails of the distribution. Busetti et al. (2015) adopt a quantile phillips-curve approach for the euro area with the aim of improving the performance of standard linear forecasting models. This paper takes a different direction with respect to the existing literature because it does not propose a standard "horse-race" exercise for competing forecasting models but rather it uses a quantile regression approach to estimate the conditional distribution of inflation forecasts and analyze its evolution over time with respect to economic conditions. Loosely speaking, this allows to study how the current state of the economy shapes the (model-based) uncertainty and perceived risks associated to future inflation with the scope of rationalizing the properties of its dynamics.

The main finding of this paper is that the distribution of euro area inflation forecasts conditional on the current state of the economy presents some asymmetries in its dynamics. Specifically, the left part of the distribution is sensitive to the deterioration of economic conditions while the right part remains relatively stable over time. In other words, downside risks to inflation vary considerably in relation to the business cycle whereas upside risks are less sensitive to economic fluctuations. This finding is robust to a large and heterogeneous set of business cycle indicators. Interestingly, some similarities are also found in the distribution of inflation forecasts obtained using the ECB Survey of Professional Forecasters (SPF). Although the evidence of an asymmetric distribution is modest, it turns out that the skewness of the distribution of inflation forecasts implied by the SPF evolves according to the current developments of economic conditions, i.e. slightly positive in good times and slightly negative in bad times.

The rest of the paper is structured as follows. Section 1.2 presents the data, the methodology and the estimated conditional distribution which is then analyzed in section 1.3. Section 1.4 performs an out-of-sample exercise and 1.5 presents some robustness checks. Section 1.6 tests some characteristics of the ECB SPF and 1.7 concludes.

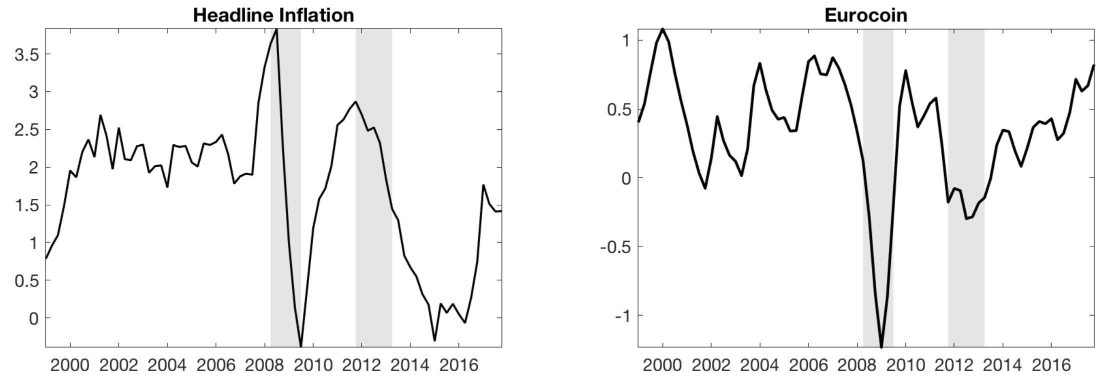
1.2 Estimating the Conditional Distribution

This section starts with the description of the data used for the empirical exercise. Then, it presents the steps required to estimate the conditional distribution of euro area inflation forecasts. First, I adopt a quantile-regression approach to capture the effect of changes in economic conditions across quantiles of the inflation distribution. Second, I show the estimated conditional distributions and outline their main characteristics.

1.2.1 Data

The evolution of the conditional distribution of inflation is analyzed with respect to the *Eurocoin*, which is a measure introduced by Altissimo et al. (2010) and can be considered as a real-time "thermometer" of the euro area economy. In practice, this index is constructed from a dynamic factor model which uses a large set of macroeconomic variables (industrial production, business surveys, financial and demand indicators and others) to extract the information that is relevant to forecast GDP. This index tracks underlying GDP growth and therefore it represents a valuable indicator of the growth rate conditions in the Euro area. Together with the year-on-year inflation rate, data are taken at quarterly frequency and the sample spans the period q1:1999-q4:2017 as illustrated in figure 1.1.

Fig. 1.1 Headline inflation (left) and Eurocoin (right)



Note: the grey shaded areas recession periods in the Euro area

The choice of using the Eurocoin rather than GDP itself is motivated by the fact that GDP is affected by a sizeable short-run component so that, for example, the beginning of a medium-run upswing cannot be distinguished from a transitory upward movement within a basically negative path. The Eurocoin tackles this problem by providing a (smooth) estimate of the *medium to long-run* component of GDP, i.e. the part of the GDP growth rate obtained by removing the fluctuations of a period shorter than or equal to one year. This feature becomes important especially in the context of forecasting future prices level. Indeed, monetary policy is targeted on price stability on the *medium* term and the assessment of the state of the economy should be based on an index at the same frequency, therefore the Eurocoin represents the best (business-cycle) candidate for this purpose. Nonetheless, evidence from other business cycle indicators (such as GDP and industrial production) will be presented in Section 1.5.

1.2.2 Quantile-regression approach

The first step is to use a quantile regression (see Koenker and Bassett (1978)) of y_t (euro area headline inflation) on x_t (the Eurocoin plus a constant). This allows the estimation of the conditional distribution in a second step as described in the next session. Equation 1.1 shows a standard quantile regression formula in which the coefficients β_α are chosen to minimize the quantile weighted absolute errors as follows

$$\beta_\alpha = \underset{\beta \in R^k}{\text{argmin}} \sum_{t=1}^{T} (\alpha \cdot \mathbb{1}(y_t \geq X_t\beta)|y_t - x_t\beta_\alpha| + (1-\alpha) \cdot \mathbb{1}(y_t < X_t\beta)|y_t - x_t\beta_\alpha|) \quad (1.1)$$

where α represents the different quantiles and $\mathbb{1}(\cdot)$ denotes the indicator function.

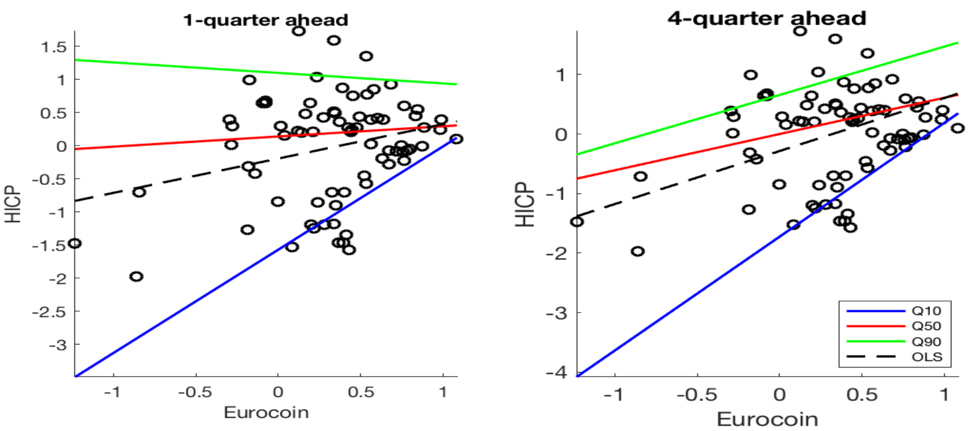
Compared to the simple linear regression model (OLS), the differences are mainly two: the minimization problem is based on the sum of *absolute* errors and not on the sum of squared errors and in addition the error terms are weighted differently according to the relative quantile (and not equally as in the OLS framework). Concretely, the use of the quantile regression approach presents several advantages. First, it allows to study the impact of the conditioning variables on different quantiles of the inflation distribution and not only on the mean as in the case of OLS. This feature is extremely important since recent periods have been characterized by two recessions and simple linear regression models might fail to capture in an appropriate way the effects of large shocks on inflation. Second, estimation and inference in a quantile-regression framework are distribution-free and therefore no strong assumptions are needed on the distribution of the inflation rate. Third, this quantile approach provides a *model*-based measure of inflation uncertainty which vary over time depending on the state of the business cycles and this is clearly something that can not be obtained via least squares.

The estimated coefficients of equation 1.1 are analyzed in two different ways: (i) their fit with the data, especially with respect to the outliers and (ii) the variation across quantiles. The first point is addressed by figure 1.2 which shows the scatter plots of the data together with the slope of 10th, 50th and 90th percentiles (blue, red and green line, respectively) and the OLS estimates (black-dashed line). These charts are informative about the ability (or inability) of the ordinary least squares in fitting the data at all quantiles of the distribution: for instance if the OLS line is parallel to the ones characterizing the other quantiles, then it indicates that a linear regression model also does a good job in capturing the relation between inflation and Eurocoin in the tails.

The evidence suggests that the OLS slope is not able to capture the relation at both tails, especially at lower quantiles. In particular, the 10th percentile slope (blue line) behave remarkably different from the other lines and this holds at both horizons. Figures 1.3 provides a more complete picture of the results by illustrating the coefficient estimates across all quantiles. First, the blue starred lines show that the coefficients are generally statistically significant (with the exception of the right tail at one-quarter ahead) [1]. Second, both graphs point out significant differences across quantiles, especially comparing the lower quantiles

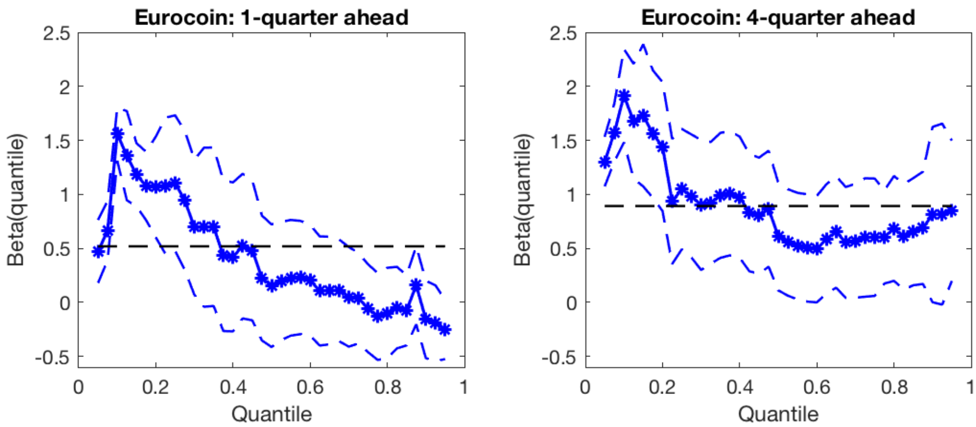
[1] The confidence bands are obtained by estimating the variance-covariance matrix as described by Greene (2008).

Fig. 1.2 Scatter plot with estimated coefficients



with the upper quantiles. Clearly, this reinforces the idea that OLS estimates (black dashed line) lack of capturing tails relation and therefore are less informative about tail risks whereas a quantile-regression approach performs well in this context and adequately fit for the purpose of this paper.

Fig. 1.3 Estimated coefficients over quantiles



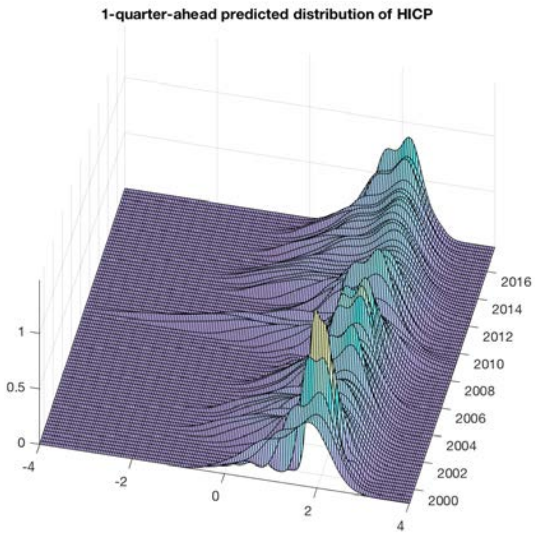
1.2.3 The Conditional Distribution

Using the estimated coefficients β_α of the quantile function of equation 1.1, the predicted distribution of inflation $Q_{y_t|x_t}$ can be easily estimated as

$$Q_{y_t|x_t}(\alpha) = x_t \beta_\alpha \,. \tag{1.2}$$

Equation 1.2 produces a set of points which are then used to obtain the predicted distribution by means of a *non-parametric* approach as the normal kernel function (e.g. D'Agostino et al. (2013))[2]. Figures 1.4 presents the estimated conditional distributions with respect to one (left) and four-quarter (right) ahead forecasts.

Fig. 1.4 Estimated distribution of Inflation forecasts



The main finding is that there is a considerable time-variation in the shape of the conditional distribution which is mainly driven by the different behaviour of the tails. Indeed, the distribution shows asymmetric moves with the left tail more sensitive to business cycle developments while the right tail remains relatively stable over time. Loosely speaking, the conditional distribution of inflation forecasts points out that the downside risk to inflation moves following business cycle conditions while the upside risk is stable over time, not increasing even when the economy is booming. As an implication, changes in inflation uncertainty are entirely driven by the left tail of the conditional distribution, therefore when an increase in the uncertainty is not generated by symmetric moves of both tails.

Interestingly, this evidence is analogous to what Adrian et al. (2016) find for the distribution of US Gross Domestic Product (GDP) conditional on financial conditions: indeed it evolves in an asymmetric way with the right tail of the distribution is quite stable over time while the left tail is strongly affected by the deterioration of financial conditions.

[2] The empirical results are robust to different specifications for the kernel function or to a simple interpolation as in Busetti et al. (2015).

1.3 Inflation Vulnerability

This section digs into the estimated conditional distribution in different ways. First, it looks at the role of conditioning the inflation distribution on economic conditions. Second, it provides a quantification of the evolution of the extreme inflation rate scenarios generated by the estimated distribution. Finally, it illustrates two appealing inflation-related policy scenarios.

1.3.1 Relative Entropy

In this section I quantify the implications of conditioning on the state of the business cycle by considering the Kullback-Leibler divergence between the conditional, $f(y)$, and the unconditional distribution, $g(y)$, of inflation forecasts[3]. As described by Adrian et al. (2016), this measure can be viewed as an indicator of vulnerability because it assesses the impact of the conditioning variable on the distribution of inflation. Analytically, this can be written as

$$\mathcal{L}_t^D(f_t, g) = \int_{-\infty}^{F^{-1}(0.5)} (log\, g(y) - log\, f_t(y))\, f_t(y) dy \tag{1.3}$$

$$\mathcal{L}_t^U(f_t, g) = \int_{F^{-1}(0.5)}^{\infty} (log\, g(y) - log\, f_t(y))\, f_t(y) dy\,. \tag{1.4}$$

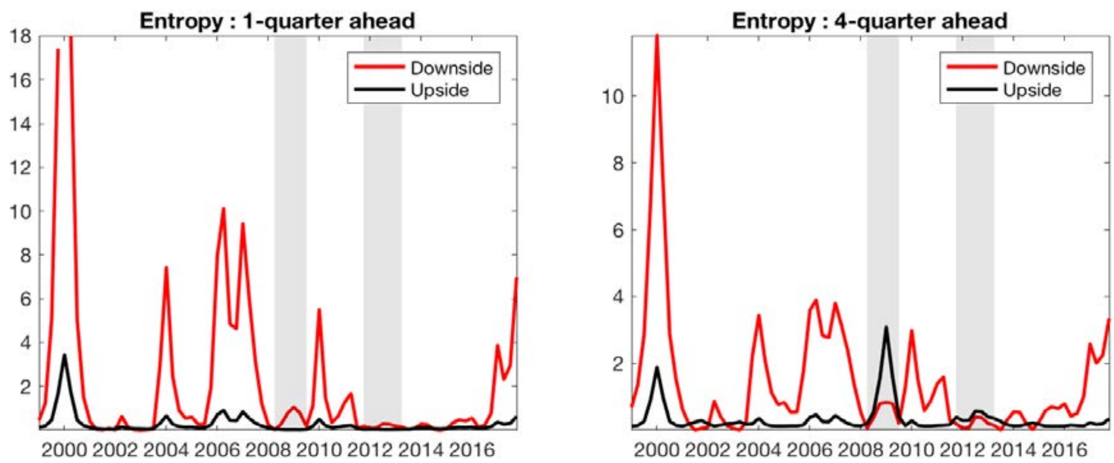
Figure 1.5 shows the relative entropy, where downside entropy $\mathcal{L}_t^D$ refers to the difference between the unconditional and conditional distribution for the left part of the distribution (namely from the first percentile to the median) and the upside entropy $\mathcal{L}_t^U$ which is the analogous for the right part of the distribution. The main evidence is that the downside entropy is significantly more volatile than the upside entropy which implies that economic conditions do not provide any information gain about the upper quantiles of the distribution, while they are extremely important for the bottom ones. In other words, the variation of the conditional distribution is mainly driven by the movements in the left tail and not from the right tail which remains relatively constant over time.

1.3.2 Expected Shortfall and Longrise

The estimated conditional distribution can also be used to quantify the expected values of extreme inflation scenarios, namely either low or high values that a forecaster might predict according to this quantile-regression model. This is implemented by considering

[3]The Kullback-Leibler divergence, also known as relative entropy and information divergence, is a measure of the non-symmetric difference between two distributions. It is generally used because differently from other measures of distance, it has a direct counterpart in the logarithmic scoring rule, which is commonly used for evaluating density forecasts (see Amisano and Giacomini (2007)).

Fig. 1.5 Downside (red) and Upside (blue) Entropy



the expected shortfall and expected longrise which are two measures commonly used in the finance literature to represent the expected return on the portfolio in the worst (or best) $\alpha\%$ of cases. In practice, these two measures can be defined as

$$SF_t = \int_0^{0.05} F_t^{-1}(\alpha)\, d\alpha \qquad LR_t = \int_{0.95}^{1} F_t^{-1}(\alpha)\, d\alpha \tag{1.5}$$

and they correspond to the integral of the probability density function at its 5th and 95th percentiles.

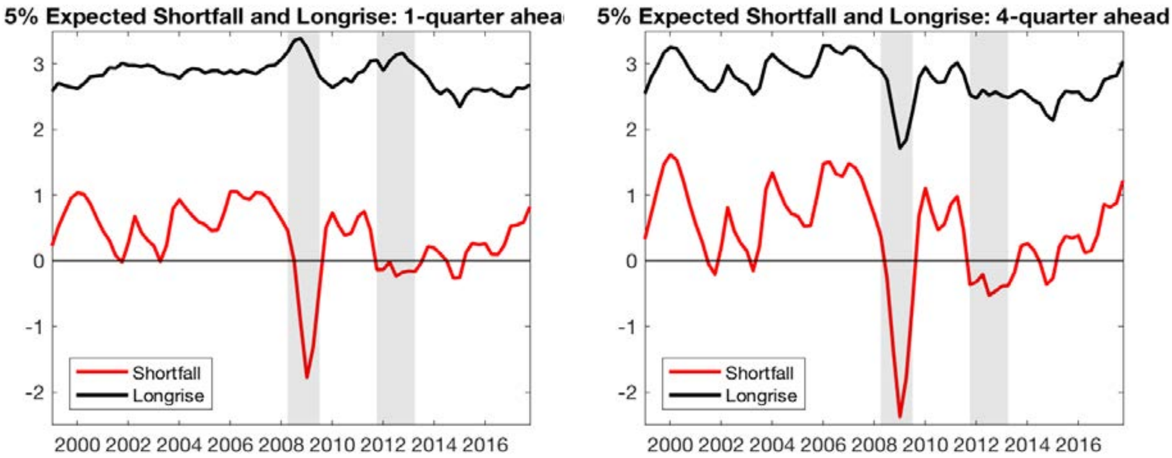
Figure 1.6 shows that he expected longrise remains stable over time fluctuating in the range between 2 and 3% while the expected shortfall is much more volatile and it can reach levels close to -2% as in the case of the Great Recession. The main message of these two figures is that the expectation of low inflation is sensitive to the fluctuations in the business cycle whereas the upside risk is less responsive to changes in economic conditions.

1.3.3 Policy-scenario probabilities

Price stability is the main target of the majority of central banks. In the Euro area, the European Central Bank (ECB) is the institution which pursues medium- and long-term price stability, with the mandate to keep headline inflation “close but below to two percent". Recent episodes as the Europe’s Double-Dip recessions and recent oil turmoils had a considerable impact on the level of prices and brought the inflation rate also in negative territory.

With this is mind, this section uses the estimated conditional distribution to propose the quantification of two relevant policy scenarios, i.e the risk of deflation (DEFL) and the

Fig. 1.6 Expected Shortfall (red) and Longrise (blue)



probability of being below the target (BT)[4]. Using the same notation of equation 1.5, this can be specified as

$$F_t^{DEFL}(0) = \int_{-\infty}^{(0)} f_t(y)dy \qquad F_t^{BT}(2) = \int_{-\infty}^{(2)} f_t(y)dy\,. \tag{1.6}$$

Figure 1.6 illustrates the quantitative results in following way: the red line represents the probability of a future inflation rate in negative territory while the black line corresponds to the probability of an inflation rate below the 2% target. Similarly, the grey and red shaded areas represent the periods in which the Euro area was below the target and in deflationary time, respectively.

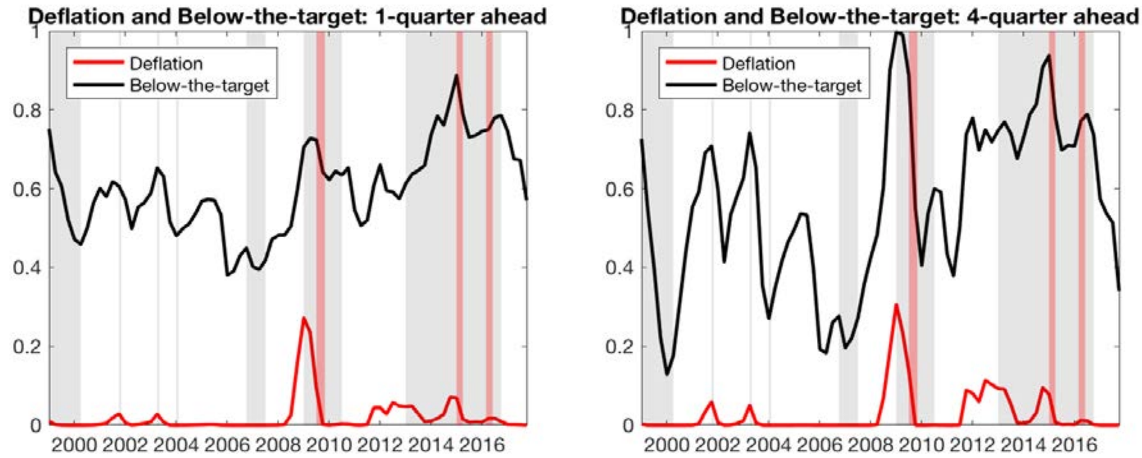
The charts suggest that on the one hand the model is more able to capture the first deflationary episode due to the fact that it was mainly generated by a big slowdown in the economic activity, while the others were more driven by the dynamics of commodities (for instance oil decreased by 66% between June 2014 and March 2016). On the other hand, inflation rates below the target are relatively well-captured by the model over the entire sample. Last comment is about the co-movement of these probabilities which is positive and consistent with the results highlighted in previous subsections.

1.4 Out-of-sample

The empirical results presented so far are obtained using an in-sample approach. Although this is extremely useful for understanding the properties of inflation forecasts, it does not fully

[4]For simplicity, the implementation of this analysis assumes that the target corresponds to a 2% inflation rate.

Fig. 1.7 Policy-scenario analysis



Note: the grey shaded areas represent the periods in which the inflation rate was below 2% while the red shaded areas correspond to deflationary periods.

replicate the true experience of a real-time forecaster. For this reason this section proposes an out-of-sample exercise to assess the forecasting ability of the model. The forecasting exercise considers the period 1999-2006 as a training sample and it evaluates the performance of the model over the period 2007-2017 using a recursive estimation procedure. The focus is on three different dimensions: (i) the predictive score, which tries to assess the forecast accuracy, (ii) the probability integral transform (PIT), that evaluates the calibration of the predictive density and (iii) the difference between the in-sample and out-of-sample tails, in order to evaluate the validity of the model in an out-of-sample context.

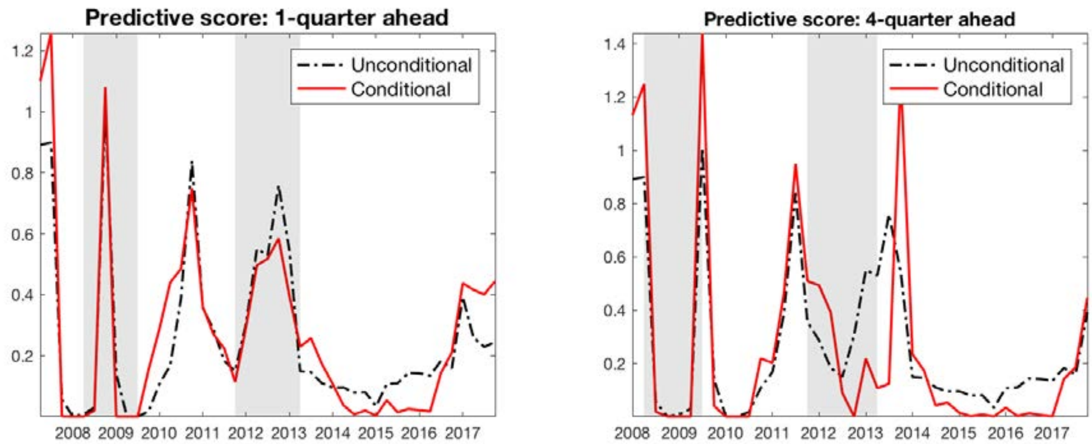
1.4.1 Predictive score

In the case of a non-parametric distribution, the predictive score corresponds to the value of the predictive density generated by the model at the realized value of inflation following the logic that the higher the score, the more accurate the model is.

Figure 1.8 shows the predictive scores for the unconditional (black dashed line) and conditional (red line) distribution. Evidence suggests that the performance of the two models is relatively similar with no evidence of superior ability between them. This finding is also supported by standard statistical test of equal accuracy (see Amisano and Giacomini (2007)) which reject the hypothesis of the significance of a constant between the score[5]. This result does not represent a concern because it is in line with the common evidence (see Atkeson and Ohanian (2001)) that inflation has become extremely hard to be predicted after the Great

[5]Results are not presented here for convenience but they are available upon request.

Fig. 1.8 Unconditional and conditional predictive scores



Moderation and simple models based only on past inflation are difficult to be outperformed.

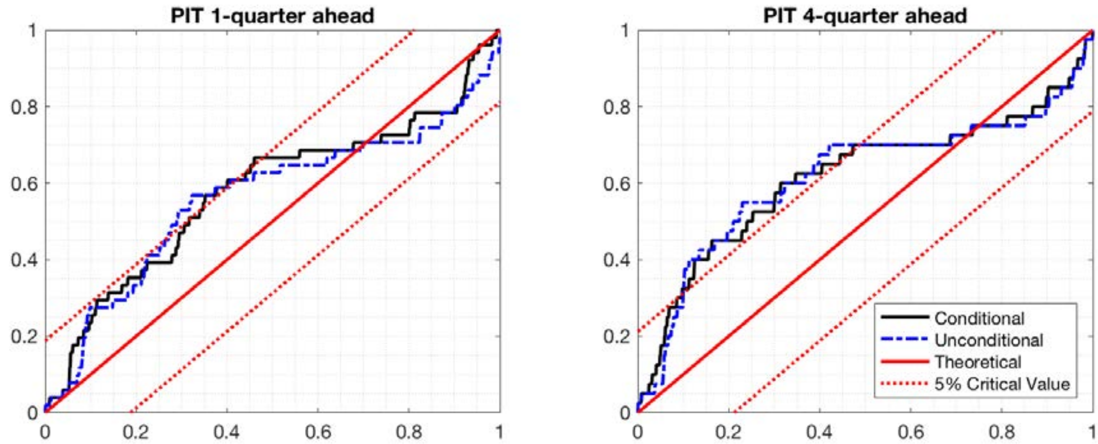
1.4.2 Probability Integral Transformation

This part analyzes the calibration of the predictive distribution by looking at the empirical cumulative distribution of the probability integral transformation (PIT), a tool introduced by Diebold et al. (1998) which is a common practice in the literature. In a nutshell, this measure indicates the percentage of observations that are below any given quantile and it evaluates whether the the empirical predictive distribution matches the true (unobserved) distribution that generates the data.

To test for correct specification of the conditional predictive density, I adopt the test proposed by Rossi and Sekhposyan (2015) which preserves the estimation error of the parameters used to construct the densities and focuses on evaluating the *absolute* performance of the model's predictive density[6]. In this context, a perfectly calibrated density should be equal to the 45-degree line, so any deviation from the bisector line suggests a bias in the predictive density. Figure 1.9 shows the results for the conditional (black line) and unconditional (blue dashed line) distribution. The main results are two: (i) the difference in terms of calibration between the two distributions is relatively small and (ii) both conditional and unconditional distributions appear relatively well-calibrated, even if they presents a left-tail bias especially at 4-quarter ahead. This left-tail bias implies that both distributions have assigned more weights to low-inflation values, indicating that the model (regardless whether it is conditioned or not on the business cycle) has been relatively pessimistic on the inflation outlook.

[6]Critical values (red dotted lines in the graphs) are obtained using the method proposed by Rossi and Sekhposyan (2015) which also provide an interesting review of the related literature.

Fig. 1.9 Probability Integral Transformation: conditional vs unconditional



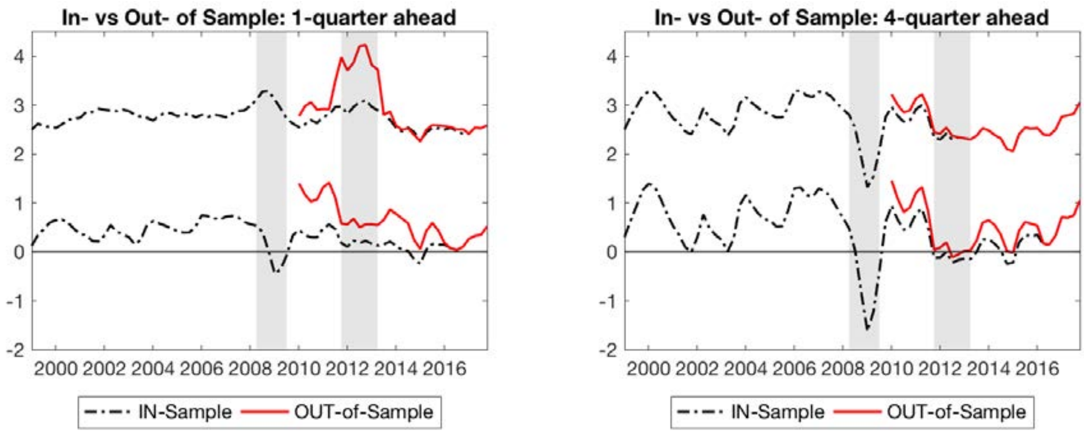
1.4.3 In-sample versus out-of-sample quantile estimates

This section sheds more light on the validity of the model in an out-of-sample context. Although this paper is focused on explaining how the business cycle affects the perceived risk of inflation (rather than improving forecasting accuracy for inflation), this *in-* versus *out*-of-sample exercise can help to understand the possible variation in the relations between the Eurocoin and inflation rate.

This exercise considers the Q1-2010 as the starting point to compare the estimated tails obtained using the two approaches[7]. Figure 1.10 presents the estimated 5th and 95th percentiles obtained using both approaches. The results can be summarized as follows. First, the difference in the 4-quarter distributions (right figure) are smaller than the 1-quarter ones (left figure) meaning that parameters' stability are less concerning at longer horizon. Second,the out-of-sample upper tail in 1-quarter differs from its in-sample counterpart in the second part of the Double-dip recession due to the phenomenon called *missing inflation* (see e.g. Jarocinski and Bobeica (2017)), namely inflation has been unexpectedly low since 2012 in contrast to what would have been predicted based on economic conditions. Overall, given the short sample in consideration, the evidence suggests that the model does a decent job in capturing tail risks also in an out-of-sample framework.

[7]Selecting an earlier date does not have much sense because the model would not have any recessions in its history, therefore I feed the model with the first big recession to see how it performs with the second one in the period 2011-2013.

Fig. 1.10 In- vs Out-of-sample estimates of tails



1.5 Robustness Checks

The results obtained so far use the Eurocoin as the indicator of the business cycle in the Euro area. Although its validity has been extensively analyzed (see Altissimo et al. (2010)), it is interesting to study whether the left asymmetry of the distribution of inflation forecasts is also robust to other business cycle measures. For this reason, I repeat the analysis of section 1.3 by conditioning on a list of standard business cycle indicators: more precisely, I consider real GDP, industrial production, ISM PMI manufacturing, unemployment rate and unemployment gap rate (see Stock and Watson (2010)), and capacity utilization[8]. Figure 1.13 and 1.14 present the distributions for the aforementioned indicators[9]. The charts confirm the main result of the paper, namely the existence of a downside risk to inflation which is generated by the deterioration of the business cycle. In some cases the evidence of a left asymmetry appears less strong than the case of the Eurocoin but this is due to the fact that the Eurocoin is obtained by including a large variety of indicators of business cycles, therefore it appears to better capture the current state of the economy.

1.6 The Vulnerability of the ECB SPF

The literature has extensively studied different characteristic of the ECB Survey of Professional Forecasters (SPF) (see e.g. Kenny et al. (2014, 2015) and Abel et al. (2016) among the others)[10]. Surprisingly, the asymmetry of these distributions has not drawn much attention.

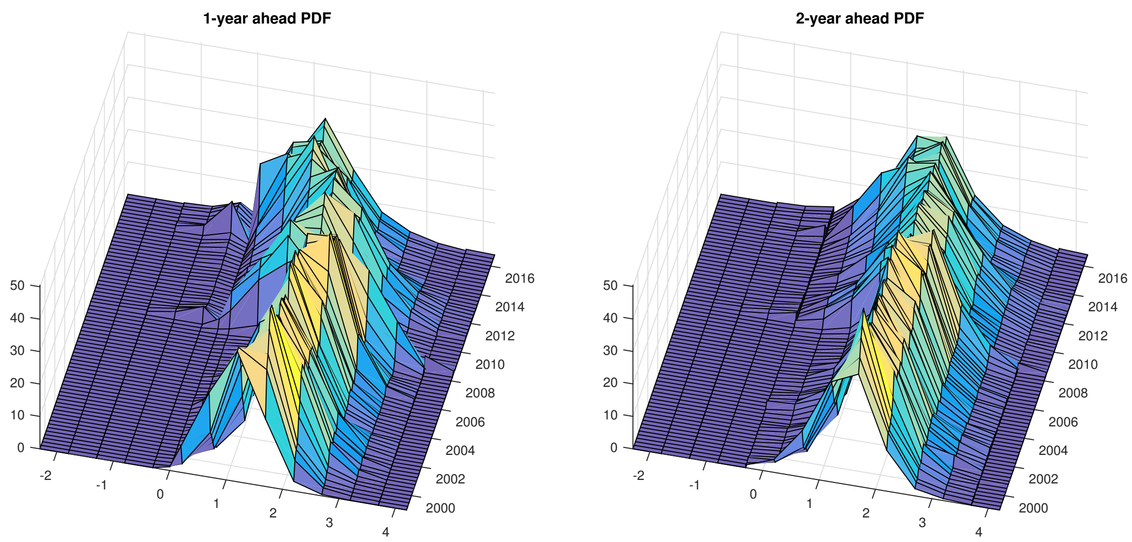
[8]A complete description of the variables is available in the Data Appendix.

[9]In this case the figures are located at the end of the paper for convenience.

[10]Garcia (2003) provide a complete description of the characteristic of this quarterly survey of expectations introduced in 1999. For interest on the data and the documentation, I refer

Indeed, empirical studies generally focus on the first two moments of the distribution and put less emphasis on higher moments. However, especially in terms of risk assessment, the skewness of the distribution helps to measure the perceived risk. For this reason, this section focuses on the asymmetry of the pdf distribution of the ECB SPF. Figure 1.11 shows the 1-year (left) and 2-year (right) probability distribution function for inflation.

Fig. 1.11 SPF probability distribution function

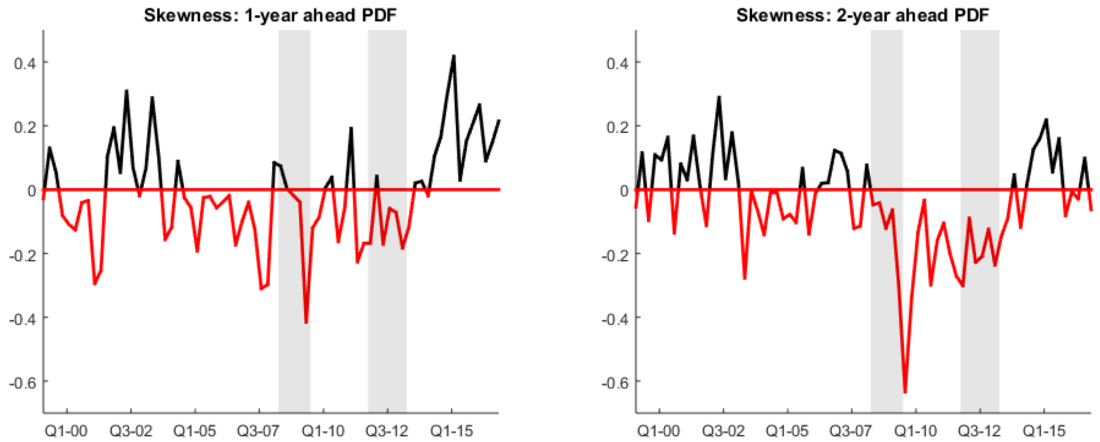


Note: The plotted distributions correspond to the aggregation of the individual probability distribution function of each forecaster.

Although these two distributions display a relatively normal-shape, a careful inspection presents some asymmetry. Indeed, as described by figure 1.12 in which I compute the asymmetry of the distribution at each point in time, there is some evidence of a non-negligible skewness whose sign appears to be related with the current state of the economy. In other words, the skewness of the distribution is generally negative when the economy experiences bad times and slightly positive during the recovery phase. Indeed, during the period of the Double-Dip recession, the asymmetry is generally negative (red line) while the last quarters of the sample are characterized by null or slightly positive asymmetry related to the optimistic expectations of future developments (black line). All in all, this evidence seems to coincide with the main finding of the previous sections.

to https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/index.en.html

Fig. 1.12 Skewness of the 1-year and 2-year ahead PDF



Note: The plotted distributions correspond to the aggregation of the individual probability distribution function of each forecaster.

1.7 Conclusion

Recent years of low-inflation rates have posed new questions in the profession (e.g. Ciccarelli and Osbat (2017)). Evaluating the risk to inflation has become of primary importance, especially for policymakers. Quantitatively, this is commonly measured by the predictive density which corresponds to an estimate of the expected probability distribution of the target variable.

To formally evaluate the risk to inflation conditional on the current state of the economy, this paper proposes a comprehensive analysis of the evolution of the distribution of inflation forecasts conditional on macroeconomic conditions. More specifically, I first adopt a quantile-regression approach to appropriately capture the effects of changes in economic conditions on all quantiles of inflation distribution. Then, I use the quantile function to estimate the conditional distribution of inflation forecasts. The main finding is that there is a significant time-variation in the shape of the distribution which is mainly due to the bottom quantiles of the distribution. In other words, the conditional distribution presents a dynamics of the downside risk which is sensitive to the current state of the economy while the upside risk remains stable over time. This evidence is generally robust to the several business cycle indicators.

This paper also performs some other interesting exercises using the estimated conditional distribution. First, it shows the implication of conditioning on the business cycle by means of the relative entropy. Second, it quantifies the probability of possible but extreme inflation outcomes and lastly it performs some relevant policy scenario analyses to study the ability of the model in capturing inflation dynamics.

Finally, a simple analysis shows an interesting result regarding the ECB Survey of Professional Forecasters. Practically, the constructed SPF distribution of future inflation presents some modest but non-negligible asymmetry which is related to the state of the economy and therefore in line with the main finding of this paper.

Appendix

The data are taken from the ECB Statistical Data Warehouse. The sample covers the period January 1999-October 2016, using quarterly observations.

- **HICP**: Euro area (changing composition) - HICP - Overall index, Monthly index, backdated, fixed euro conversion rate used for weights, European Central Bank, Working day and seasonally adjusted;

- **Eurocoin**: Euro area (changing composition), Centre for Economic Policy Research and Banca d'Italia, Coincident indicator of business cycle, based on quarterly changes in cyclical component of the GDP, see Altissimo et al. (2010);

  For the robustness exercise:

- **real GDP** Gross domestic product at market prices - Euro area 19 (fixed composition) Chain linked volume (rebased), Non transformed data, Calendar and seasonally adjusted data;

- **Industrial production**: Euro area 19 (fixed composition) - Industrial Production Index, Total Industry (excluding construction); Working day and seasonally adjusted;

- **ISM PMI Manufacturing**: Euro area 19 (fixed composition), Markit, Manufacturing - output, Total, Seasonally adjusted, not working day adjusted;

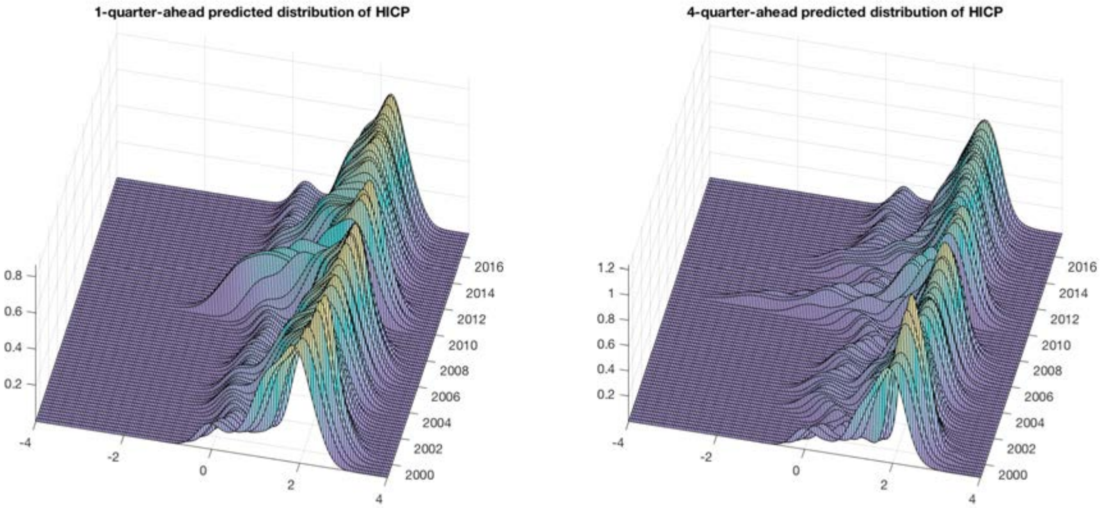
- **Unemployment rate**: Euro area 19 (fixed composition) - Standardised unemployment, Rate, Seasonally adjusted, not working day adjusted, percentage of civilian workforce;

- **Unemployment gap rate**: this measure corresponds to the difference between the current unemployment rate and the minimum unemployment rate over the last 12 quarters (see Stock and Watson (2010));

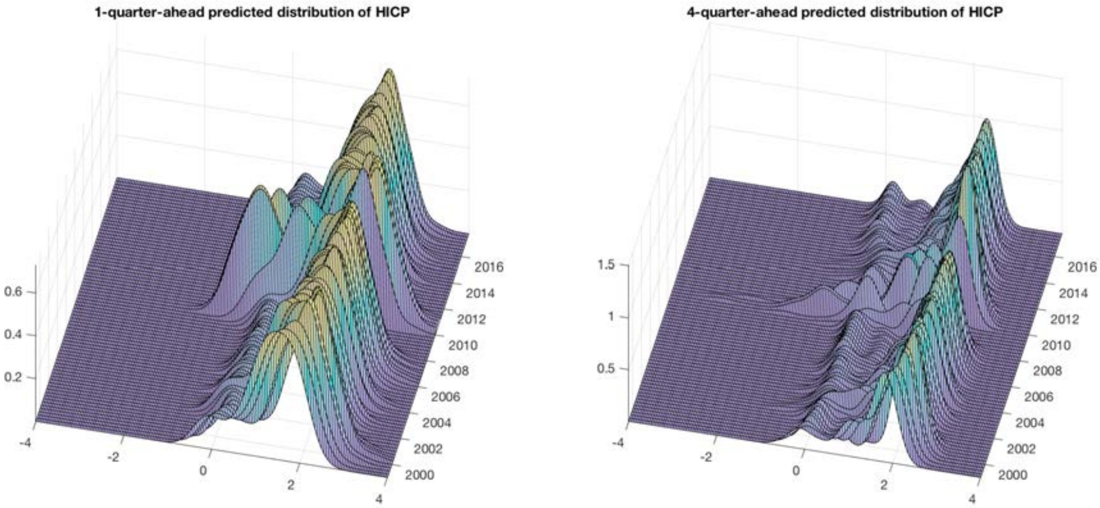
- **Capacity utilization**: Euro area 19 (fixed composition), EU Commission, DG-ECFIN, Industry survey - current level of capacity utilization, Seasonally adjusted, Percentages.

Fig. 1.13 Estimated distribution of Inflation forecasts

a) Real GDP growth rate



b) Industrial Production



c) ISM PMI Manufacturing

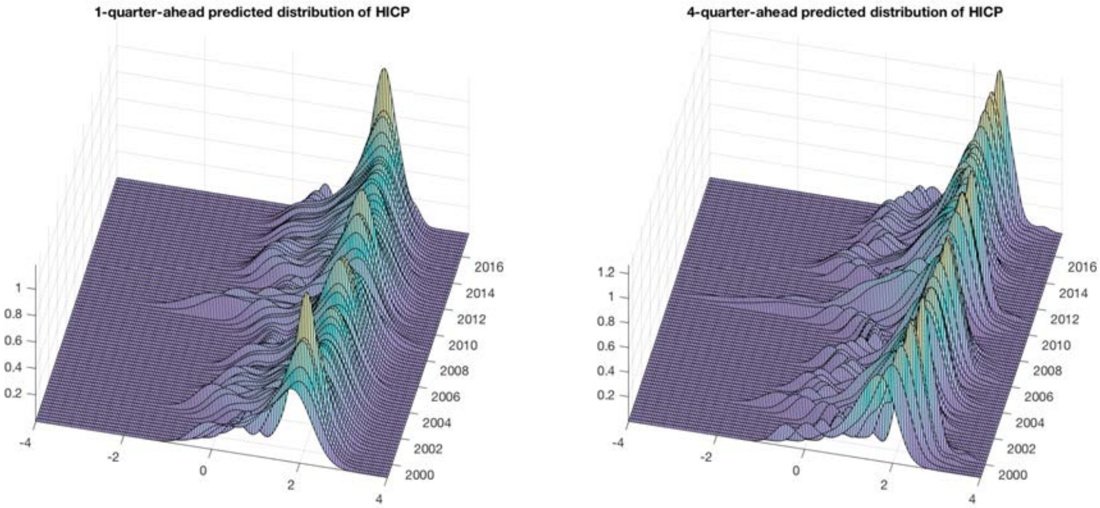
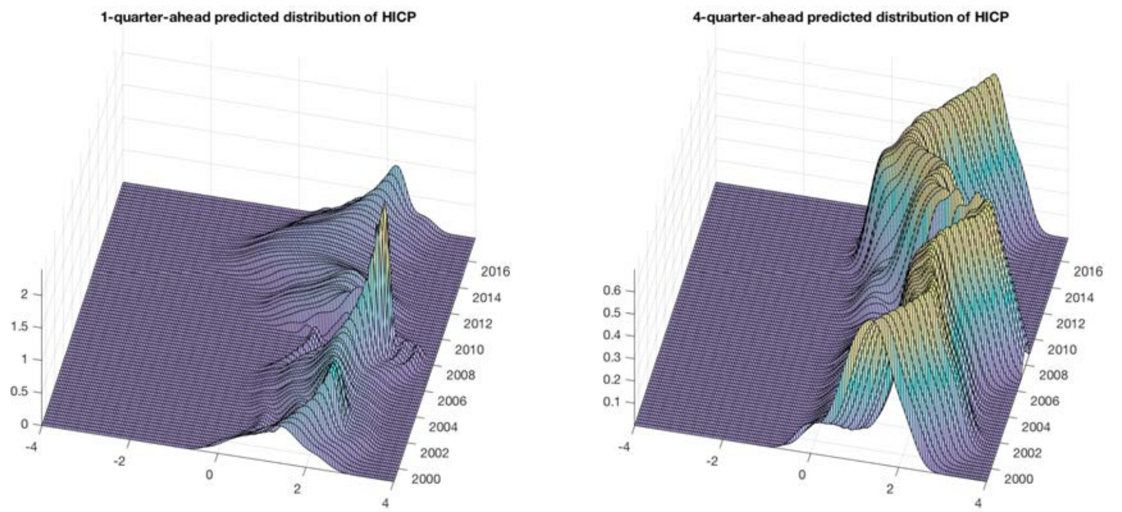
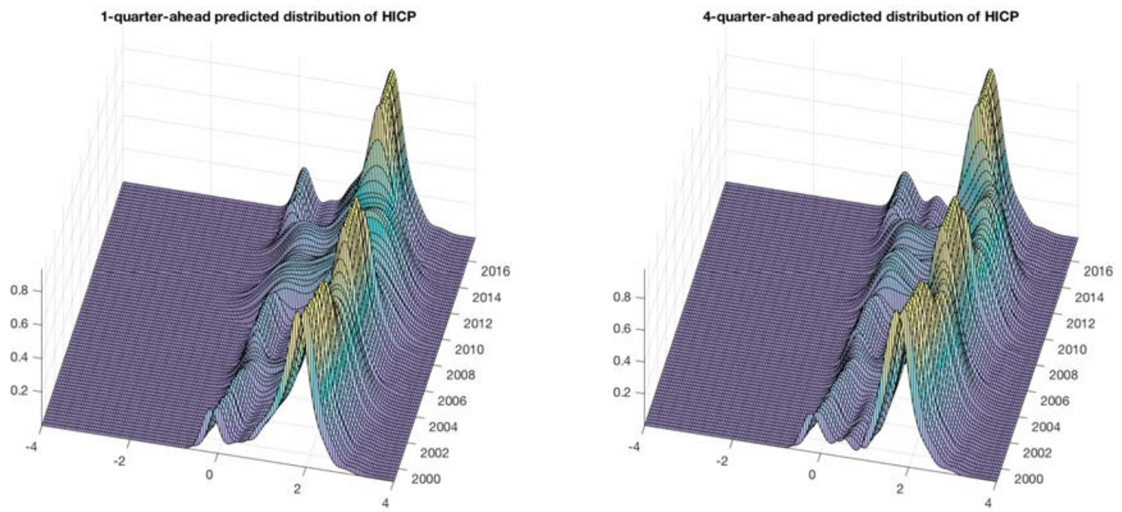


Fig. 1.14 Estimated distribution of Inflation forecasts

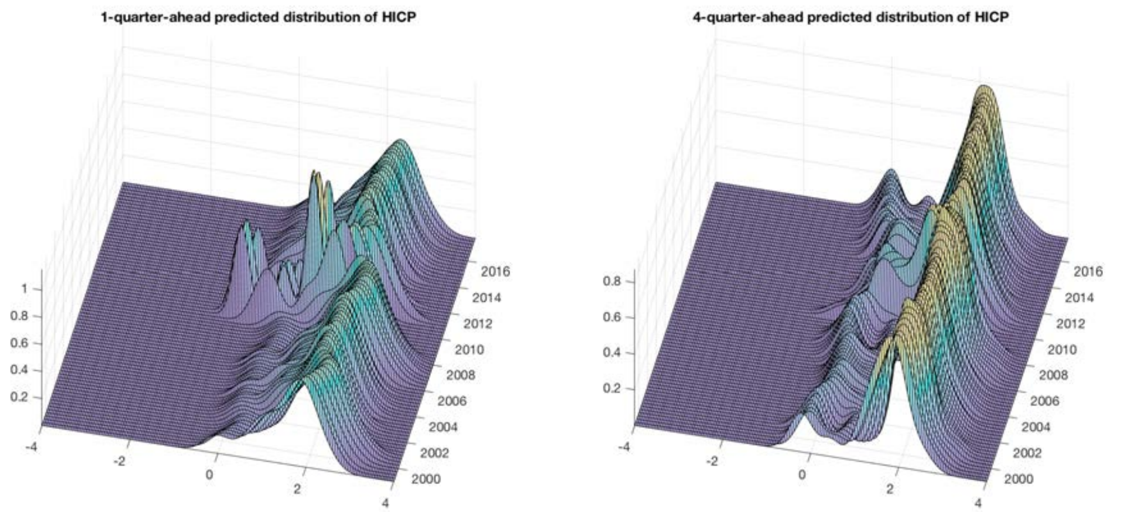
d) Unemployment rate



e) Unemployment gap rate (see Stock and Watson (2010))



f) Capacity Utilization



Chapter 2

Nowcasting Italian GDP using multivariate mixed-frequency models

Abstract

Exploiting timely and high frequency data are standard practices in the nowcasting literature. This paper proposes and compares two multivariate econometric models that take advantages of both characteristics to predict Italian GDP, namely a dynamic factor model and a mixed-frequency Bayesian VAR. A pseudo out-of-sample exercise shows three main results: (i) both models considerably outperform a standard univariate benchmark; (ii) the dynamic factor model turns out to be more reliable close to the release date while the mixed-frequency BVAR appears superior with an incomplete information set; (iii) the overall forecasting superiority of the dynamic factor model is mainly driven by its ability in capturing the severity of large recession episodes.

JEL classification: C32, C38, C53, E37

Keywords: Nowcasting, Italian GDP, Dynamic Factor Model, Mixed-frequency BVAR

2.1 Introduction

The literature on nowcasting the growth rate of gross domestic product (GDP) has extensively developed over the recent years. Following the seminal paper by Giannone et al. (2008), the dynamic factor model (DFM hereafter) has become the standard workhorse to provide accurate forecasts not only for GDP but also for other macroeconomic variables. This model is characterized by the ability of capturing the co-movement among a large set of macroeconomic variables, even at different frequency, but at the same time guaranteeing parsimony. This econometric framework has been shown to work well for advanced economies, like France (Barhoumi et al. (2009)), Germany (Antipa et al. (2012)), Ireland (D'Agostino et al. (2012)), Norway (Aastveit et al. (2014)) and the Euro Area (Marcellino et al. (2016)), but also for emerging economies (Dahlhaus et al. (2017) for BRIC countries).

Similarly, there is a recent strand of the nowcasting literature that developed the so-called mixed-frequency Bayesian VAR (henceforth MFBVAR), which allows to handle a large number of times series observed at different frequencies. In this case, rather than extracting common factors as in the case of DFM, this model combines the flexibility of a state-space representation with Bayesian shrinkage techniques to handle a large number of variables. The seminal contribution by Schorfheide and Song (2015) and subsequent works by Carriero et al. (2015) and Brave et al. (2016) are some examples that illustrate not only the validity of this type of models for forecasting purposes, but also the advantage of including mixed-frequency variables in standard quarterly frequency Bayesian VARs.

However, a quantitative comparison between these two models has not been shown yet. To this end, this paper aims to fill this gap by proposing a dynamic factor model and a mixed-frequency BVAR to nowcast Italian GDP growth rate and then comparing their relative performance to identify their main advantages. In other words, this paper presents a "horse-race" forecasting exercise between the two most-performing "horses" in the nowcasting literature using Italian GDP as an empirical case. With respect to an existing study on Italian GDP, this paper differs from Aprigliano et al. (2017) where they focus on the predictive power of payment system data in a mixed-frequency DFM while this paper considers a wide and heterogeneous set of macroeconomic variables. In addition, the adoption of a MFBVAR for a country different from the United States represents an interesting case of study.

The main results can be summarized in few points. First, both models considerably outperform a standard univariate benchmark demonstrating their well-known forecasting ability also in the context of Italian GDP. Second, an exercise that mimics the actual flow of data releases shows that the MFBVAR appears more accurate when the information set is incomplete but the DFM improves considerably at the end of the forecasting period. Third, the overall forecasting superiority of the DFM comes from the ability of capturing the downturns of the economy, not only in terms of promptness but especially in terms of severity of the crisis.

2.2 Data

This paper considers a set of variables that aims to exploit three important features: (i) correlation with GDP, (ii) short delay in the publication lag (also known as timeliness) and (iii) higher frequency than GDP (i.e. monthly versus quarterly). The selected dataset is composed by sixteen variables which combines soft and hard indicators to capture different signals from the economy. The choice of using a medium-size dataset is motivated by the following reasons: first, as highlighted by Bańbura et al. (2013), rich models are not necessarily used to improve the forecasting accuracy but rather to evaluate and interpret any (new) significant information that may have an impact on the nowcast. Similarly, Boivin and Ng (2006) argue that large specification does not necessarily help to improve the accuracy because factor models would tend to extract a more *noisy* signals with the possible consequence of a deterioration of the forecasting ability. This is also confirmed by Alvarez et al. (2016) who show in a Montecarlo exercise that the performance of models with 10-30 variables is as accurate as models with a hundred variables. On the same line, Poncela and Ruiz (2016) demonstrate that factor extraction remains robust to small-medium specifications, therefore not requiring a large cross-section.

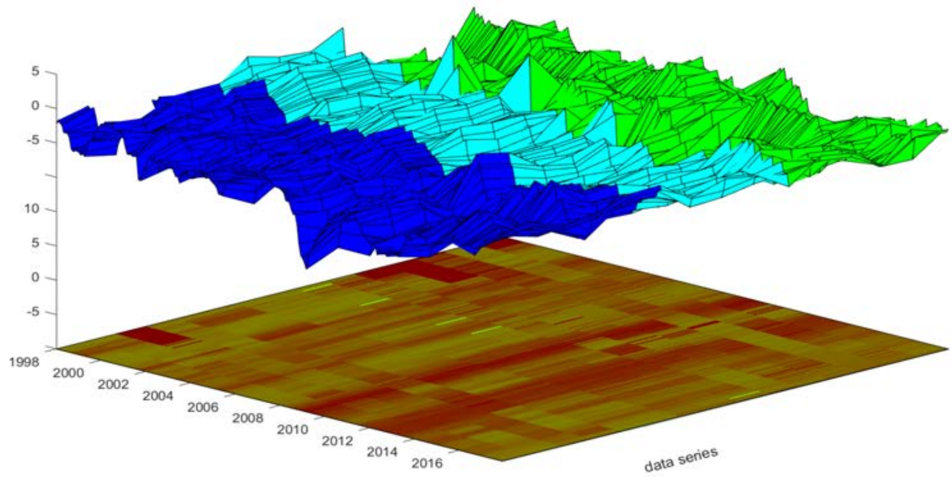
Table 2.1 Data description

Description	Timing	Frequency	Source	Starting date	Transformation
Business Confidence Climate	-3	M	ISTAT	1998	Index
Business Survey: Economy in next 3months	-3	M	ISTAT	1998	Index
Business Survey: Production in next 3months	-3	M	ISTAT	2000	Index
Consumer Confidence	-3	M	ISTAT	2000	Index
Manufacturing PMI	5	M	Markit	1998	Index
New Orders (Manufacturing) PMI	5	M	Markit	1998	Index
Services (New business) PMI	5	M	Markit	1998	Index
Services (Business activity) PMI	5	M	Markit	1998	Index
Composite PMI Output	5	M	Markit	1998	Index
New passenger Cars	5	M	ISTAT	1998	MoM%
New orders	20	M	ISTAT	1998	YoY %
Industry Turnover	20	M	ISTAT	2000	YoY %
Cassa Integrazione Ordinaria	20	M	INPS	1998	3MMA
Retail Sales	20	M	ISTAT	1998	MoM %
Industrial Production	40	M	ISTAT	1998	YoY %
GDP	**45**	**Q**	**ISTAT**	**1998**	**QoQ %**

This paper considers a set of soft and hard indicators whose characteristics are described in Table 2.1. The block of soft indicators includes both PMI (from Markit) and confidence indicators from the Italian Statistical Office (ISTAT) with the purpose of capturing different and (possibly) heterogeneous signals from soft data which are generally more reactive to changes in the dynamics of the business cycle. Similarly, the block of hard data is composed by standard but extremely important real activity indicators which tend to be strongly correlated with the real GDP. With respect to the of publication lag (second column), also in the Italian case soft data are characterized by shorter publication delays than hard data.

Data described in Table 2.1 are illustrated in Figure 2.1. The sample covers the period from January 1998 to December 2017 when possible and all indicators are taken in growth rates when appropriate to achieve stationarity. The heatmap reported in the bottom of the figure, which shows for each variable values above (yellow) or below (red) their relative mean, sheds light on the fact that there is a strong co-movement among macroeconomic variables.

Fig. 2.1 Data



Note: In the upper graph, dark blue areas represent PMI data, light blue ones describe "Climi di Fiducia" and light green indicates "real activity indicators". The heat map in the bottom horizontal plane describes the observations above (yellow) and below (red) their relative mean in which the intensity of the color is function of the size of the deviation.

2.3 The Econometric Models

2.3.1 The dynamic factor model

The methodology follows the approach introduced by Giannone et al. (2008). Denoting with $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{n,t})'$ a n-dimensional vector of stationary monthly variables, x_t is assumed to have this representation:

$$x_t = \mu_t + \Lambda f_t + \epsilon_t \tag{2.1}$$

where f_t is a $r \times 1$ vector of unobserved common factors, Λ is a $n \times r$ matrix of factor loadings and ϵ_t is a vector of idiosyncratic components. The common factors have zero mean and μ represents the unconditional mean. As in Bańbura and Modugno (2014), ϵ_t is considered to follow an AR(1) process

$$\epsilon_{i,t} = \alpha_i \epsilon_{i,t-1} + e_{i,t}\,, \quad e_{i,t} \sim i.i.dN(0, \sigma_i^2) \tag{2.2}$$

with $E(e_{i,t}e_{j,s}) = 0$ for $i \neq j$. The common factors are modeled as a stationary vector autoregressive process (VAR) of order p (in this case equal to 2):

$$f_t = A_1 f_{t-1} + A_2 f_{t-2} + \ldots + A_p f_{t-p} + u_t\,, \quad u_{i,t} \sim i.i.dN(0, Q) \tag{2.3}$$

where $A_1, A_2, \ldots A_p$ are $r \times r$ matrices of autoregressive coefficients. This dynamic factor model can be easily cast into a state-space representation, as described in the Appendix. For a more detailed description, see Giannone et al. (2008) and Bańbura and Modugno (2014).

Quarterly series

The inclusion of a quarterly variable as GDP in a model specified with monthly indicators is implemented using the approach introduced by Mariano and Murasawa (2003) which allows to express a quarterly variable in terms of a partially observed monthly indicator. In this case, the quarterly level of GDP, denoted as GDP_t^Q, can be expressed as the sum of its unobserved monthly counterparts (GDP_t^M):

$$GDP_T^Q = GDP_t^M + GDP_{t-1}^M + GDP_{t-2}^M\,, \quad t = 3, 6, 9 \ldots \tag{2.4}$$

Assuming $Y_t^Q = 100 \times log(GDP_t^Q)$ and $Y_t^M = 100 \times GDP_t^M$, then the monthly growth rate of GDP corresponds to $y_t = \Delta Y_t^M$. This leads to represent GDP as

$$y_t = \mu_Q + \Lambda_Q f_t + \epsilon_t^Q \tag{2.5}$$

and the link between the observed and the unobserved counterpart of GDP is described as

$$y_t^Q = \begin{cases} Y_t^Q - Y_{t-3}^Q, & t = 3, 6, 9, 12 \\ \text{unobserved} & \text{otherwise.} \end{cases} \tag{2.6}$$

Assigning quarterly observations to the last month of the quarter, Mariano and Murasawa (2003)'s method produces

$$
\begin{aligned}
y_t^Q = Y_t^Q - Y_{t-3}^Q &= (Y_t^M + Y_{t-1}^M + Y_{t-2}^M) - (Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M) \\
&= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}
\end{aligned}
\tag{2.7}
$$

whose representation imposes a set of restrictions on the factor loadings for the monthly GDP growth rate.

Estimation

The literature proposes different methods to estimate dynamic factor models[1]. This paper follows Bańbura and Modugno (2014) which applies quasi-maximum likelihood estimation by using the Expectations Maximization (EM) algorithm and the Kalman smoother. This approach guarantees three main advantages: (i) efficiency in small sample, (ii) (easily) dealing with an arbitrary pattern of data availability and (iii) it also allows to impose restrictions to extract the factors.

This method involves two steps. First, it calculates the expectation of the log-likelihood conditional on the parameter estimated in the previous iteration. Second, it re-estimates the parameters using the expected log-likelihood from the previous step. These two steps are iterated until the convergence of the log-likelihood is achieved. Bańbura and Modugno (2014) offers a more exhaustive technical description.

2.3.2 MF-BVAR

This paper follows the econometric framework developed by Schorfheide and Song (2015) which assumes that the economy evolves at monthly frequency according to the following VAR(p) representation:

$$
x_t = \Phi_1 x_{t-1} + \ldots + \Phi_p x_{t-p} + \Phi_c + u_t \quad u_t \sim iidN(0, \Sigma) \tag{2.8}
$$

where x_t is a $n \times 1$ vector of macroeconomic variables specified at different frequencies, namely $x_t = [x'_{m,t} x'_{q,t}]$.

[1] Poncela and Ruiz (2016) provide a complete overview of the existing methodologies and their relative performance with small and large datasets.

Define $z_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]$ and $\Phi = [\Phi_0, \ldots \Phi_p, \Phi_c]$. Then, the corresponding VAR(1) companion form equals to

$$z_t = F_1(\Phi)z_{t-1} + F_c(\Phi) + v_t \quad v_t \sim iidN(0, \Omega(\Sigma)) \tag{2.9}$$

which describes the state-transition equation of the MF-BVAR. The characterization of the measurement equation requires few steps. First, denote the actual observations as

$$y_{m,t} = x_{m,t} \tag{2.10}$$

and assuming that the underlying VAR has at least three lags, the three-month average of $x_{q,t}$ is described as

$$\tilde{y}_{q,t} = \frac{1}{3}(x_{q,t} + x_{q,t-1} + x_{q,t-2}) = \Lambda_{qz} z_t. \tag{2.11}$$

Let $M_{q,t}$ be a selection matrix that equals the identity matrix if t corresponds to the last month of a quarter and empty otherwise, then it can be written

$$y_{q,t} = M_{q,t}\tilde{y}_{q,t} = M_{q,t}\Lambda_{qz} z_t. \tag{2.12}$$

Let $y_{m,t}$ be the subset of monthly variables for which t observations are reported after period T and $M_{m,t}$ be a sequence of selection matrices, then it can be specified as

$$y_{m,t} = M_{m,t} x_{m,t} \tag{2.13}$$

which leads to compress all previous equations to a more compact expression for the measurement equation

$$y_t = M_t \Lambda_t z_t \quad t = 1, \ldots T. \tag{2.14}$$

Bayesian estimation

The estimation procedure adopted in this paper follows Schorfheide and Song (2015) which is based on the use of the Gibbs sampler to generate draws for the parameters (Φ, Σ) and for the latent states $(Z_{0:T})$ from the posterior distribution. This is obtained by factorizing the joint distribution of data, latent variables and parameters, namely

$$p(Y_{1:T}, Z_{0,T}, \Phi, \Sigma | Y_{-p+1:0}, \lambda) = p((Y_{1:T}, Z_{0,T})p(Z_{1,T}|z_0, \Phi, \Sigma)p(z_0|Y_{-p+1:0})p(\Phi, \Sigma|\lambda) \tag{2.15}$$

into the following representation

$$\begin{aligned} p(\Phi, \Sigma | Z_{0:T}, Y_{-p+1:T}) &\propto \quad p(Z_{1:T}|z_0, \Phi, \Sigma)p(\Phi, \Sigma|\lambda) \\ p(Z_{0:T}|\Phi, \Sigma, Y_{-p+1:T}) &\propto \quad p(Y_{1:T}|Z_{1:T})p(Z_{1:T}|z_0, \Phi, \Sigma)p(z_0|Y_{-p+1}) \end{aligned} \tag{2.16}$$

which allows to the Gibbs sampler to iterate over the two conditional posterior distributions in (2.16) following Carter and Kohn (1994). The choices of the prior (standard Minnesota prior plus dummy observations), of the hyper-parameters and of the number of lags (which is set equal to three) are based on a calibration exercise reported in the Appendix.

2.4 Empirical Results

This section is composed by three parts. Section 2.4.1 performs a *pseudo* out-of-sample exercise to evaluate the accuracy of both models in predicting output assuming the information set before the GDP data release, i.e. at $t + 45$. Section 2.4.2 shows the evolution of the accuracy accuracy mimicking the actual flow of data availability. Finally, Section 2.4.3 tests whether the quantitative results holds also using a *qualitative* criterion by focusing on the directional accuracy. All three exercises are implemented estimating both models with a recursive approach and the evaluation covers twelve years over the period Q1:2006-Q4:2017. The forecast horizon regards either the current quarter (h=0), also known as the nowcast or the forecast for the next quarter (h=1).

2.4.1 Forecast Accuracy

The quantitative performance is evaluated using the mean squared error (MSE). The results are presented in relative terms with respect to a benchmark model corresponding to an autoregressive model of order one, which is parsimonious but behaves relatively well in the context of predicting Italian GDP. This corresponds to the formula

$$\text{Relative MSE}_i = \frac{(1/T)\sum_t(\hat{y}_t^i - y_t)^2}{(1/T)\sum_t(y_t^{AR} - y_t)^2} \tag{2.17}$$

where the forecasts $\hat{y}_t^i$ correspond to the DFM or the MFBVAR and y_t identifies the actual value of GDP. The evaluation considers the results for the nowcast of the current period ($h = 0$) and for the forecast of the next period ($h = 1$) over the entire sample and also over two sub-sample splits, each of them characterized by one recession, though with different magnitudes. The numbers in in Table 2.2 should be read as follows: numbers below one indicates superior forecasting ability of the model with respect to the benchmark and the percentage gain (or loss) of using such a model is quantified by the difference with respect to one.

Table 2.2 outlines two main results. On the one hand, both models significantly outperform the benchmark in the range of 30-35% over the full sample in the context of nowcast (h=0).

Table 2.2 Out-of-sample performance using the Mean Squared error

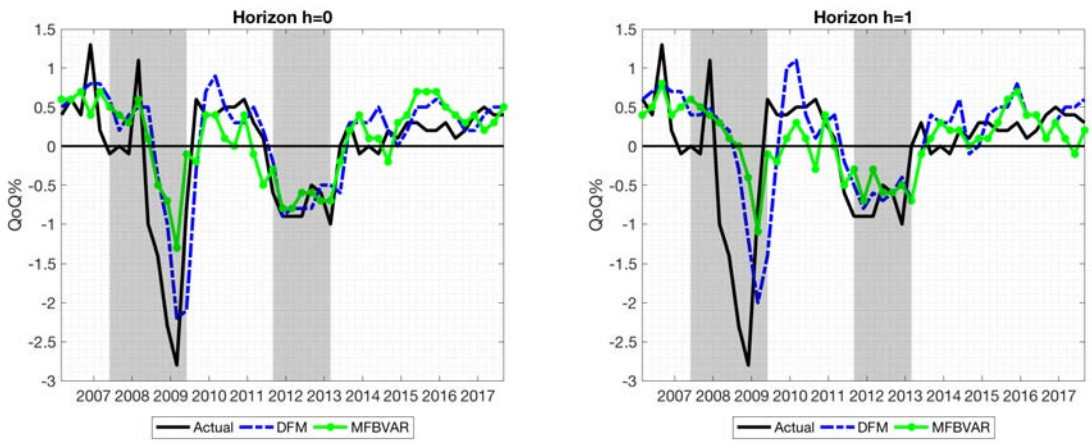
	Nowcast h=0			Forecast h=1		
Model	**Full**	**2006-11**	**2012-17**	**Full**	**2006-11**	**2012-17**
AR(1)	1	1	1	1	1	1
DFM	**0.65**	**0.61**	1.00	**0.72**	**0.68**	1.26
MFBVAR	**0.69**	**0.66**	**0.91**	**0.70**	**0.66**	1.16

Note: bold numbers indicates superior accuracy with respect to the benchmark AR(1) model.

This accuracy gain comes mostly from the first part of the sample while the advantages decrease after 2012. On the other hand, the DFM shows a modest forecasting superiority with respect to the MFBVAR that also in this case is obtained in the first part of the sample. Results are qualitatively similar for the one-quarter ahead exercise, with an overall gain in the range of 30% that is affected by the relatively poor performance in the second period of the sample.

Figure 2.2 clarifies these findings in a graphical way. Indeed, the DFM (blue dashed line) is not only more timely in identifying the slowdown of the economy in 2008 but also better captures the severity of that recession. Contrarily, the MFBVAR (green dotted line) does not fully identify the considerable slowdown of the economy, predicting a one percent decrease versus the actual drop by minus three percent. Interestingly, the 2012 Sovereign Debt crisis, which appears less severe compared to the previous one, is properly captured by both models in terms of promptness and magnitude.

Fig. 2.2 Illustration of the forecast accuracy



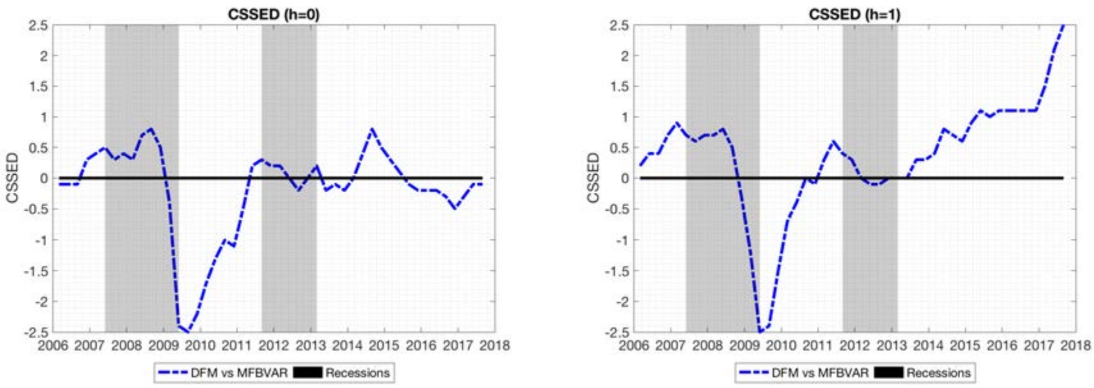
The comparison between these two models is further investigated using the cumulative sum of squared prediction error difference (CSSED) which represents a measure for tracking the evolution in accuracy of point forecasts (see Pettenuzzo and Timmermann (2017)) among others). Analytically, this measure can be described as

$$CSSED_{i,t} = \sum_{t=t_0}^{T} (e^2_{DFM,t} - e^2_{MFBVAR,t}) \tag{2.18}$$

where $e^2_{DFM,t}$ and $e^2_{MFBVAR,t}$ indicate the squared forecast errors of the dynamic factor model and the mixed-frequency BVAR, respectively. Values of this measure above (below) zero indicates that the DFM performs relatively worse (better) than the MFBVAR, while positive (negative) changes in the slope represent a decrease (increase) in relative performance of DFM.

Figure 2.3 illustrates this comparison. First, the CSSED becomes considerably negative during 2009 confirming the result that the DFM captures more appropriately the magnitude of the recession. However, this measure rapidly converges to zero in the aftermath of the crisis because the MFBVAR tracks the recovery more precisely instead of over-predicting the growth acceleration as the DFM. Starting from 2011, the scenario is different because both models perform similarly with the only exception at the end of the sample when the DFM recovers some advantage. Second, this pattern characterizes also the one-quarter ahead horizon, with the only exception that the DFM becomes remarkably less accurate in predicting GDP in the last part of the sample.

Fig. 2.3 Evolution of the relative performance between models

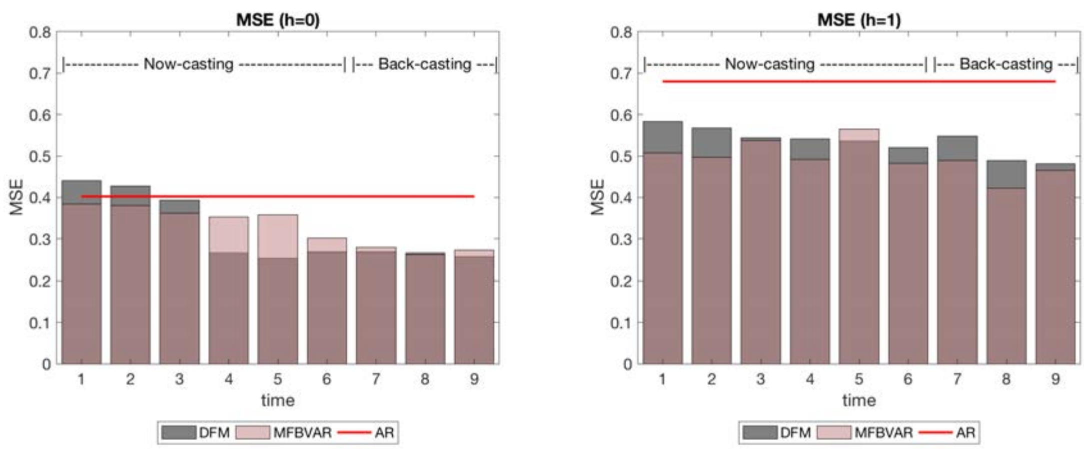


2.4.2 Evolution in the accuracy

The results obtained in previous section are obtained considering the information set as of the week of the GDP release, i.e. around $t + 45$ where t represents the end of the target period. However, policymakers consider certainly important to have accurate predictions even before the final GDP release. For this reason, this part proposes a *pseudo* real-time simulation, in which the latest available vintage of data is appropriately cut period by period being careful to replicate the missing values' pattern available within each forecasting period. In other words, this exercise aims to mimic the timing pattern described in Table 2.1, cutting the forecasting period in nine forecast rounds with an interval of one every two weeks, therefore augmenting the information set with the macroeconomic releases occurred over the last fifteen days. For instance, considering the first quarter of the year, one is interested to assess the evolution of the model accuracy from the beginning of January to the mid of April, which corresponds to the period of GDP release, with a fifteen-day interval. The forecasts calculated from January to the end of March are considered "*now-cast*" because they refer to the quarter in consideration while the predictions from the first of April onward belong to the "*back-cast*" period.

Figure 2.4 illustrates a comparison of the evolution of the accuracy between the two models over the specified nine forecasting rounds. Looking at the left panel (nowcast), the two models behave slightly different. On the one hand, the MFBVAR is more accurate at the beginning of the quarter and its accuracy gradually improves within the forecasting period. On the other hand, the DFM is outperformed at the beginning of the forecasting period but improves considerably starting from the fourth forecasting round which differs from the previous ones by including the GDP release of the previous quarter. Then, the performance at the end of the forecasting round reflects the results reported in Table 2.2. The scenario turns out to be different for the right panel (h=1) in which both models slightly improves their accuracy as new information is included, but there are no major changes between forecasting rounds.

Fig. 2.4 Evolution of MSE over forecasting rounds



Note: each bar represents the mean squared error obtained estimating each model over the entire sample using the missing-data pattern available in the specified forecasting round.

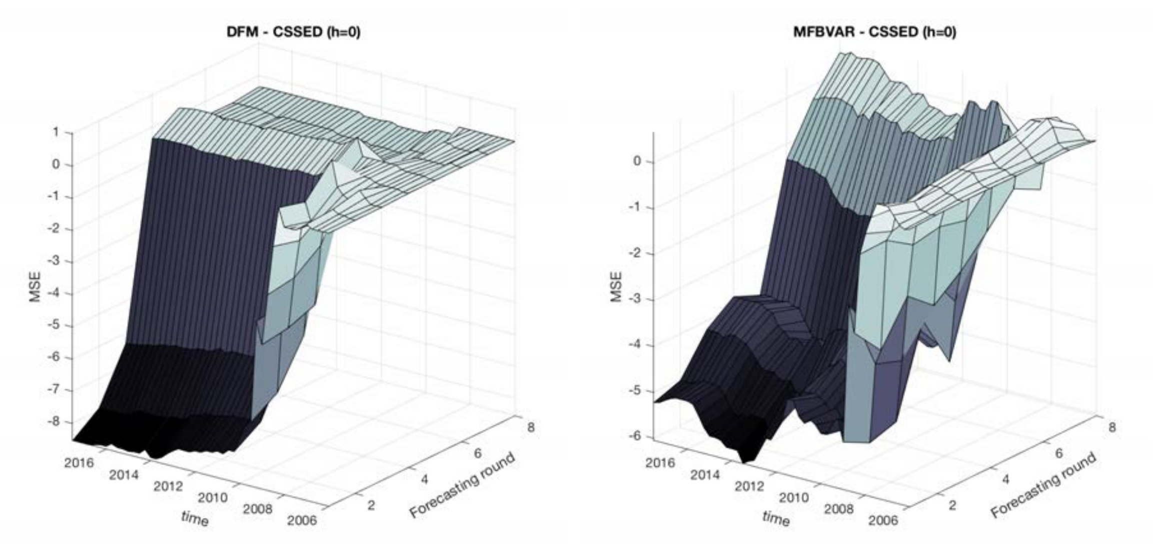
Next, the cumulative sum of squared error difference is used from a different perspective. Specifically, it is used to test whether the expansion of the information set played a significant role in different periods. Indeed, new macroeconomic releases might be more relevant before recessions because they could (and in principle should) signal the upcoming economic slowdown. This is quantitatively investigated looking at differences across forecasting rounds, namely

$$CSSED_{i,t,j} = \sum_{t=t_0}^{T}(e^2_{i,t,pre-release} - e^2_{i,t,j-period}) \tag{2.19}$$

where $i = \{DFM, MFBVAR\}$, t represents the time and $j = 1, \ldots 8$ describes the forecasting round in consideration. The last forecasting round before the GDP release is considered as the benchmark model to gauge the difference in the performance with the forecasting round including full information set.

Figure 2.5, which illustrates the results in a three-dimension framework, should be read as follows: negative numbers indicate that the model estimated in the forecasting round j performs worse than the one estimated with full information set and the intensity of the color describes the accuracy gain, with the rule that the darker the larger. The left panel, which

Fig. 2.5 Evolution of MSE over time and across forecasting rounds



Note: darker (lighter) colors represents worse (better) forecasting accuracy measured in terms of mean squared error.

considers the DFM, shows that the accuracy gain between the benchmark forecasting round and the first three comes mainly from the 2009 recession. This result suggests that the model improved considerably between the third and the fourth round, which corresponds to the period in which the previous-quarter GDP release is included. In other words, the model has quickly incorporated the severity of the crisis in its projections. The evidence appears different for the MFBVAR (right panel): the accuracy gain among forecast rounds improves more smoothly across forecasting rounds, even if the the 2009 period remains an important driver of the accuracy.

2.4.3 Directional Accuracy

In the business of forecasting the evolution of business cycles, good predictions regards not only the quantitative side, but also *qualitative* one. In other words, in some cases detecting the *direction* of the output growth can be as important as capturing the size. For this reason, this section replicates previous analyses by using a directional measure of accuracy as the mean directional accuracy (MDA), which consists of

$$\text{MDA}_i = \frac{100}{T} \sum_{t=1}^{T} [\text{sign}(\hat{y}_t^i - y_{t-1}) = \text{sign}(y_t - y_{t-1})] \tag{2.20}$$

which is reported in percentage terms and the results are expressed in terms of the percentage gain with respect to the benchmark model.
Table 2.3 shows that capturing the direction of output growth rate appears extremely challenging, given that the accuracy of the benchmark model turn out to be weak. Overall, both models outperform the benchmark, especially for the nowcasting exercise, but the percentage gain remains modest, i.e. in the range of 10-18%. However, this result suggests that both models are far from being reliable because the overall accuracy reaches values around 50%, which can be achieved with a simple naive approach.

Table 2.3 Out-of-sample performance using the Mean Directional Accuracy

	Nowcast h=0			**Forecast h=1**		
Model	**Full**	**2006-11**	**2012-17**	**Full**	**2006-11**	**2012-17**
AR(1)	34.78	37.5	31.81	43.48	33.33	54.54
DFM	**13.04**	**16.67**	**9.09**	-4.34	**8.33**	-18.18
MFBVAR	**17.39**	**16.67**	**18.18**	**6.52**	**16.67**	-4.54

Note: bold numbers indicates superior qualitative accuracy with respect to the benchmark model. The logic corresponds to the more positive the more accurate.

Note that these qualitative results should be considered with caution because the specification of both models, namely the choice of the hyper-parameters and lags selection and the variable selection, are not taken to maximize the directional accuracy, but oriented toward the quantitative accuracy (see the calibration exercise in the Appendix).

2.5 Conclusions

The nowcasting literature has considerably developed over the last years. The well-known dynamic factor model proposed by Giannone et al. (2008) has a new competitor as the mixed-frequency Bayesian VARs. This paper proposes two versions of these econometric models that use a medium-size set of hard and soft indicators to predict Italian GDP.

The empirical findings of a pseudo real-time exercise can be summarized in three main points. First, both models considerably outperform a standard univariate benchmark demonstrating their well-known forecasting ability also for Italian GDP. Second, an exercise that mimics the actual flow of data releases show that the MFBVAR is more accurate when the information set is incomplete but the DFM becomes more accurate at the end of the forecasting period. Third, the overall forecasting superiority of the DFM comes from the ability of capturing the downturns of the economy, not only in terms of promptness but especially in terms of severity of the crises.

Appendix

State-space representation of the DFM

- Measurement equation

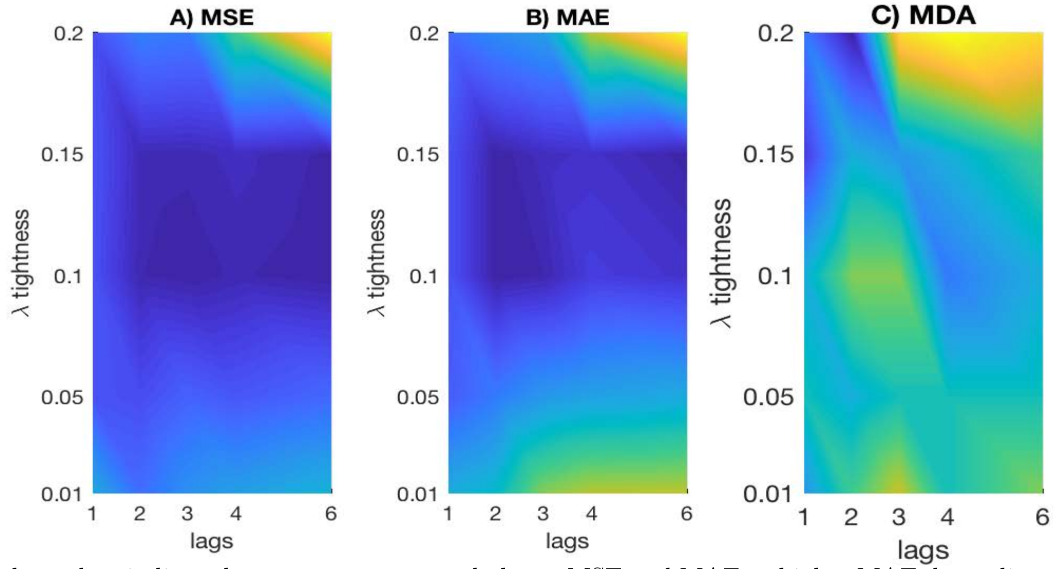
$$\begin{pmatrix} x_t \\ y_t^Q \end{pmatrix} = \begin{pmatrix} \Lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q & 1 & 2 & 3 & 2 & 1 \end{pmatrix} \times \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \varepsilon_{t-3} \\ \varepsilon_{t-4} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \xi_t^Q \end{pmatrix} \quad (2.21)$$

- State equation

$$\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{pmatrix} = \begin{pmatrix} A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ f_{t-5} \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \\ \varepsilon_{t-5}^Q \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ 0 \\ 0 \\ 0 \\ \varepsilon_t^Q \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (2.22)$$

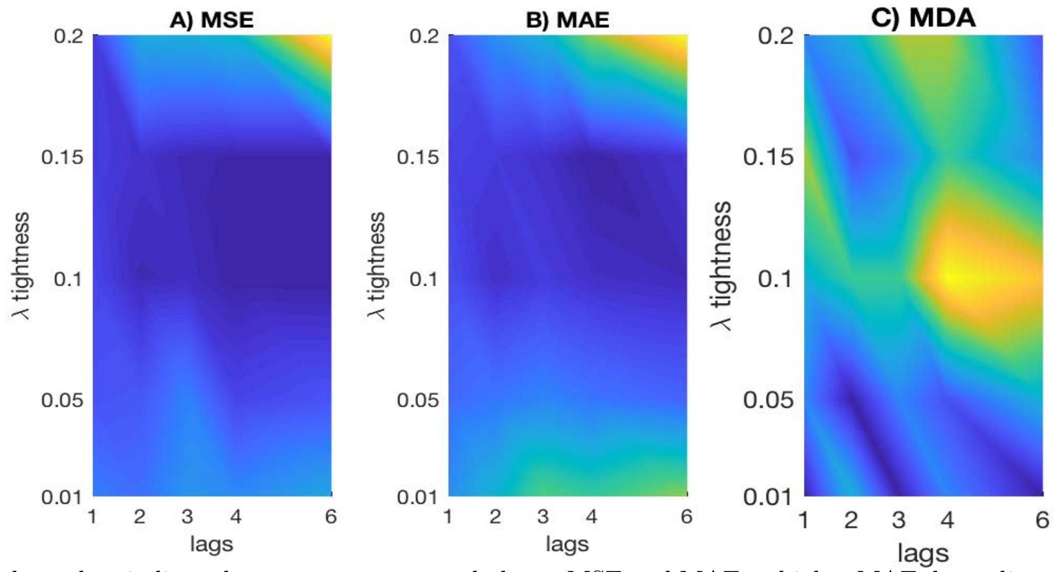
MFBVAR calibration: lags and hyper-parameters selection

Fig. 2.6 Calibration of the MF-BVAR - nowcast



Note: darker colors indicate better accuracy, namely lower MSE and MAE or higher MAE depending on the corresponding evaluation criterion. Each point in the graphs is obtained after estimating the model with the Gibbs sampling adopting the selected lags and tightness criterion.

Fig. 2.7 Calibration of the MF-BVAR - 1-quarter ahead



Note: darker colors indicate better accuracy, namely lower MSE and MAE or higher MAE depending on the corresponding evaluation criterion. Each point in the graphs is obtained after estimating the model with the Gibbs sampling adopting the selected lags and tightness criterion.

Chapter 3

Is Anything Predictable in Market-Based Surprises?

(joint with Luca Brugnolini and Antonello D'Agostino)

Abstract

We analyze two of the most important data releases that impact the US financial market, namely the change in the level of non-farm payroll employment (NFP) and the manufacturing index published by the Institute for Supply Management (ISM). We examine the unexpected component of these two, as measured by the deviation of the actual release from the Bloomberg Consensus. We label it as the market surprise, and we investigate whether its structure is partially predictable and in which cases. Secondly, we use high-frequency data on the eurodollar as a laboratory to study the effect of these surprises. We show in a regression framework that although the in-sample fit is sufficiently good, the performance deteriorates in an out-of-sample setting because a naive model can hardly be beaten in a sixty-minute window after the release. Finally, we demonstrate that under certain circumstances there is some structure that can be exploited and we provide a framework to take advantages of it.

JEL: E44, G14, G15

Keywords: macroeconomic surprises, out-of-sample forecasting, high-frequency data

3.1 Introduction

Chrysopoeia has always been the most ambitious abilities in humankind history. Nevertheless, nobody has ever discovered its correct formulation. Even in recent times, under the label of researchers, modern alchemists have attempted to create gold out of algorithms investigating and predicting the functioning of financial markets. Many have tried, more had failed. In this article, we do not aim to succeed. Notwithstanding, we exert a strenuous effort in providing a firm ground against what market agents would be able to spot a profitable trade around the major macroeconomic releases. In addressing this topic, we intend to fill a significant gap in the macro-financial literature. To the best of our knowledge, the existing empirical studies which evaluates the effects of market-based surprises on asset prices focus on analyzing the average reaction using an *in*-sample framework. Some relevant examples are Ehrmann and Fratzscher (2005), Faust et al. (2007) and Andersen et al. (2007) which analyze the reactions of a large set of assets (mainly exchange rates) to a variety of monetary policy and data announcements. The natural question is: are market-based surprise responses predictable in an *out*-of-sample framework?

In this paper, we try to answer this question by providing a comprehensive forecasting exercise. We summarize our main findings as follows. First, we start by analyzing two of the most important macroeconomic surprises obtained from the releases of Non-Farm Payroll and ISM Manufacturing index (henceforth NFP and ISM), which are known for their relevance and popularity as business cycle indicators for the US economy. Interestingly, we show that the hypothesis of market efficiency, which implies that markets are rational and agents use the entire set of available information, might fail also for these well-tracked indicators due to the presence of bias and anchoring in markets' forecasts.

Secondly, we focus on predicting the market reaction to these macroeconomic surprises. In our exercise, we use the eurodollar spot rate to test whether we can predict price movements after a macroeconomic releases[1]. In line with the empirical results obtained in the literature, we show that a simple regression model augmented with non-linear terms performs relatively well in the context of a *in*-sample framework. More specifically, dummies capturing possible size, sign and time effects are strongly significant for both variables and over different samples. However, the role played by these non-linearities appears weaker when considered in an

[1] We select the eurodollar as our target variable for two main reasons: first, markets are open before and after both ISM and NFP releases, differently for example from the US stock market (e.g., S&P500) which opens at the same time of the NFP release. This feature allows us to proxy the effect of a macroeconomic release in a short window around the event. Secondly, exchange rates are commonly considered the asset with the strongest correlation with data announcements because they are mostly driven by fundamental factors, see for example Li et al. (2015) which also provide an exhaustive review of the literature on the high-frequency analysis of macro releases on foreign exchange markets.

out-of-sample framework. Indeed, a naive model based on holding constant the first-minute response easily outperforms any model used in the in-sample analysis.

Overall, the contribution of this paper is twofold. On the one hand, we use an out-of-sample approach to show the poor predictive performance of different models highlighted in the in-sample literature. To explain this predictive gap, we look more in details at the data to provide some possible reasons and we study whether there could be some components that can be systematically predicted. Interestingly, the relation between the sign of the surprise and the sign of the asset response goes often against the theory and this can generate sizeable forecast errors, especially in an out-of-sample approach. On the other hand, we build an appropriate framework to analyze the relation between market-based surprises and the markets' response. In particular, we illustrate how to model the surprises in different regression models and predict the minute-by-minute response of an asset in a short window around the data release which generated the surprise. Practically, this procedure can be used to study any combination of market-based surprises and assets available at high frequency and therefore it represents a useful tool to help market-operators in taking trading decisions.

The rest of the paper is structured as follows. Section 3.2 describes the data. Section 3.3 shows some properties characterizing the macroeconomic surprises whose effects are then analyzed in section 3.4 and 3.5 with an in-sample and out-of-sample approach, respectively. Section 3.6 draws some conclusions.

3.2 Data description

This section describes the data used for our empirical analysis. We start presenting two critical macroeconomic indicators from which we construct the market-based surprises, i.e, the US ISM Manufacturing index and the Non-farm Payroll. Then, we illustrate the asset price quoted at a tick level on which we evaluate the effects of these surprises.

3.2.1 Why ISM Manufacturing index and Non-farm Payroll

We focus on the releases of the survey on economic activity in the manufacturing sector produced by the INstitute for Supply Management (ISM) and the US non-farm-payroll (NFP) published by the Bureau of the National Statistics (BNS), which are commonly considered two of the most significant US monthly indicators in terms of market reactions (Gürkaynak et al., 2005; Swanson and Williams, 2014). Specifically,the ISM Manufacturing is constructed from a survey of more than three hundred (manufacturing) firms, and it is considered one of most *timely* measure of the business cycle. Instead, the latter indicator measures the change in the level of persons employed in the non-farm sector (excluding government sector and no-profit organization) in the month preceding the release. As a measure of job creation, it is commonly seen as a proxy for US economic activity. Our sample covers the period February 1999-September 2017. Both variables are released early in the following month (generally in the first week), and this explains, in addition to their intrinsic value, why these are considered crucial information by market operators and policymakers. These two indicators are commonly included in any study focused on surprise analysis, both to study their corresponding effects on the markets or to construct a multi-indicator surprise and uncertainty indexes (e.g., Scotti 2016).

To measure market expectations, we use data from the Bloomberg Survey of Forecasters which includes a large set of investment banks and economic institutions actively operating in the financial markets. In previous work, Tagliabracci (2018) shows that the median of this survey represents a reliable measure to predict these two variables, being able to outperform other combinations of the forecasts included in this panel. As we will see in the next section, macroeconomic surprises are simply the standardized difference between the actual release and the Bloomberg median forecast.

Figure 3.1 illustrates the actual series (solid black line) with the corresponding median forecasts (blue dashed line) derived from the set of individual forecasts (grey triangles) included in the Bloomberg Survey. These charts highlight two main points. On the one hand, the median forecast tracks the evolution of these variable fairly well, demonstrating the reliability of the Bloomberg Survey as an accurate indicator. On the other hand, there is evidence of a non-negligible level of disagreement among forecasters surrounding the median forecast, which might be an important source of markets' reaction generated by data releases.

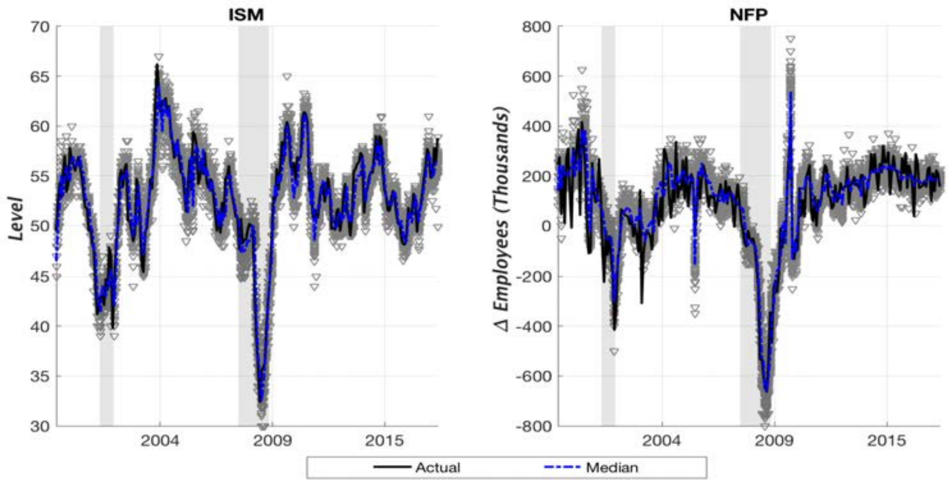
3.2.2 High-frequency data

We use tick level data from Thomson Reuters Tick History from February 1999 to September 2017. We use the eurodollar spot rate to asses market reaction. In particular, we take the average of the bid and ask quotes offered. Data are filtered from misquotes and aggregated at minute level using the last offered price in a minute (closing price). In case there are no transactions in a minute, we carry forward the last minute available. With respect to a backward carrying procedure, carrying forward avoid including information coming from the future in the past.

3.3 Forecast errors and macroeconomic surprises

The behavior of professional forecasters has been extensively analyzed in previous works. The literature based on survey data highlighted two critical facts: (i) forecasts produced by experts exhibit predictable errors and (ii) forecasters tend to disagree. This evidence

Fig. 3.1 actual series and median forecast.



Note: NFP numbers are expressed in terms of change in the level of employees in thousands unit while ISM is a diffusion index scaled at fifty. Therefore a number above (below) fifty is perceived as a proxy for good (bad) economic conditions.

characterizes different survey data as it has been found for instance by previous works on the ECB Survey of Professional Forecasters (Andrade and Bihan, 2013) and the Consensus Economics Survey (Dovern, 2013). Following this literature, this section first studies whether this evidence also applies to the Bloomberg Survey of Forecasts by looking at the properties of the forecast errors and the disagreement concerning ISM and NFP forecasts. Then, we characterize the forecast errors once we standardize them to derive the so-called *surprises*.

3.3.1 Predictability

Previous studies rationalize the idea of predictability of the forecast errors mainly with two approaches. On the one hand, Sims (2003) proposes a *noisy information* model in which agents regularly update their information but have an *imperfect* access to it at each period due to a limited processing capacity or costly access to information. In other words, the average (or the median) forecast incorporates only a fraction of new information, which makes the forecast error predictable with respect to the (perfect) information available ex-post to the econometrician. On the other hand, Mankiw and Reis (2002) provide an explanation based on a *sticky information* model in which agents have *perfect* information but they update their information set *infrequently*, generating the same implications of the noisy information framework[2].

[2] Coibion and Gorodnichenko (2012) consider both approaches and use aggregated survey data to examine and disentangle these two channels, finding mixed evidence.

To investigate these hypotheses, we propose some standard market efficiency (or rationality) tests aiming at discovering the presence of some predictive bias and market efficiency as in Andrade and Bihan (2013) and Scotti (2016), and for anchoring, i.e., forecasts leaning towards past releases, as in Campbell and Sharpe (2009)[3].

Table 3.1 presents the results together with some descriptive statistics. The first column

Table 3.1 forecast errors, descriptives and predictability.

Average forecast errors, (e_t)	NFP	ISM
1. Descriptive Statistics		
Mean(e)	-16.48	0.16
$\sqrt{Mean(e^2)}$	80.09	1.91
Skewness	-0.44	0.08
2. Rationality Tests		
a) Bias: $e_t = \alpha + u_t$		
$\hat{\alpha}$	-16.48(-3.46)	0.16(1.30)
b) Efficiency: $e_t = \alpha + \gamma e_{-1} + u_t$		
$\hat{\gamma}$	0.04(0.52)	-0.04(-0.59)
c) Market Efficiency: $e_t = \alpha + \beta \hat{F}_t + u_t$		
$\hat{\beta}$	-0.01(-0.38)	-0.02(-0.80)
d) Anchor: $e_t = \alpha + \theta(\hat{F}_t - A_{t-1}) + u_t$		
$\hat{\theta}$	0.01(0.21)	0.20(1.72)

Note: the upper panel shows the descriptive statistics of the forecast errors for the sample period covering 1999 to 2017, such as mean, standard deviation and skeweness. The bottom panel reports the results of the four different methodologies to test for rational behaviour, as described in the text. In the table, e_t represents the forecast errors, $\hat{F}_t$ the median forecast, α is a constant, and u_t corresponds to the residual errors. T-statistics are reported in parenthesis.

shows that NFP forecasts exhibit a significantly negative bias, implying that agents tend to overpredict the change in the level of employment[4]. Differently, the bias is positive but considerably weaker for ISM forecasts which also appear less volatile and skewed. Tests on efficiency suggest that there is no evidence of a relation between the forecast errors with both lags and market information, while there is evidence of anchoring in the context of ISM, namely the median forecast tend to lean towards the previous release. We find these results quite interesting. For instance Félix et al. (2018) finds that the outcome of these tests should be generally related to the degree of attention attached to these indicators, in the sense that

[3]More precisely, anchoring implies that the forecast $F_t = \lambda E(A_t) + (1 - \lambda)A$, where A is the anchor. Since $e_t = A_t - F_t$, then we can rewrite the equation as $e_t = \alpha + \theta(F_t - A) + u_t$ which is the specification estimated and shown in Table 3.1.

[4]The Appendix 3.6 shows that this negative bias comes mainly from a strong bias affecting some specific month as May and August. The reasons behind monthly seasonality remain still an object of investigation.

rationality should disappear for less popular variables, given that forecasters put less effort on the prediction of these variables and use more information from previous releases. For this reason, given the popularity of these indicators, the results go partially against this theory because bias and anchoring seem to persist also in the case of two important indicators as NFP and ISM.

3.3.2 Disagreement

Disagreement among forecasters can be interpreted as a consequence of agents' imperfect information, namely agents can have different information sets due to heterogeneous perceptions of the reality. The seminal paper by Ottaviani and Sorensen (2006) proposes two different theories to rationalize agents' behavior: on the one hand, forecasters might undervalue private information and be biased towards the prior mean for *reputational* concerns leading to decrease the level of disagreement. On the other hand, forecasters engaged in a forecasting contest might *strategically* misreport their information by putting more weight on private signals compared to what they would do in an optimal-setting, hence differentiating with respect to the other forecasters and increasing the divergence among forecasts.

To characterize the degree of disagreement, we use the cross-sectional standard deviation of forecasts at each release date as in Pericoli and Veronese (2015), and described in equation (3.1)

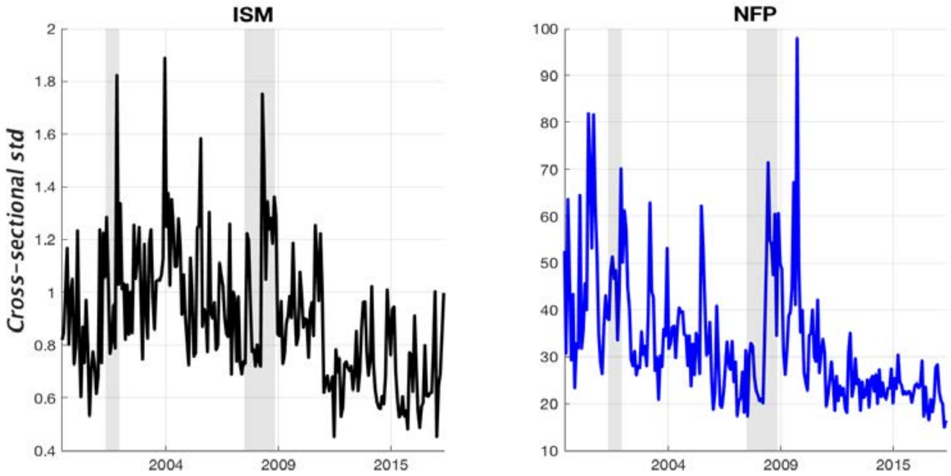
$$\sigma_t = \sqrt{(1/n_t)\sum_{i=1}^{N}(\hat{F}^x_{i,t} - \hat{F}^x_{mean,t})^2} \tag{3.1}$$

where $\hat{F}^x_{i,t}$ represents the individual forecast at time t, and $\hat{F}^x_{mean,t}$ the corresponding mean[5]. Figure 3.2 illustrates the level of disagreement for the two variables. It turns out that disagreement evolves, with a considerable increase during recession periods, as in the case of ISM, or in the periods after the economic slowdown as for NFP. Overall, the degree of disagreement for these two variables is positively correlated indicating that the uncertainty surrounding the state of the business cycle appears similar.

We then investigate possible drivers of the evolution of the level of disagreement. Specifically, the main idea is to test whether disagreement increases with respect to the amplitude of the shocks hitting the economy, which can be described with measures such as the squared of the change in the level of the variable $(\Delta X_{t-1})^2$, the squared of the previous forecast error $(e_{t-1})^2$ and the squared of the change of the Bloomberg median forecast $(\Delta\hat{F}_t)^2$. Table 3.2 shows that the estimated coefficients are all significantly positive for both variables meaning that disagreement is positively correlated with new information about the state of

[5]The mean and the median forecast look quantitatively very similar, therefore results are robust to the use of both measures. We use the mean to follow the corresponding literature.

Fig. 3.2 disagreement among forecasters.



Note: the black line represents the cross-sectional standard deviation of forecasts over time. Grey shaded areas represent NBER recession dates.

Table 3.2 disagreement among forecasters.

Disagreement, σ	NFP	ISM
1. Descriptive Statistics		
Mean(σ)	32.93	0.90
std.dev (σ)	13.61	0.25
$\rho_\sigma(1)$	0.62	0.51
2. Tests		
Regression, σ on		
a) $(\Delta X_{t-1})^2$	0.00013^{***}	0.0143^{***}
b) $(e_{t-1})^2$	0.000542^{***}	0.020^{***}
c) $(\Delta\hat{F}_t)^2$	0.000131^{***}	0.0165^{***}

Note: the table shows the disagreement for NFP and ISM respectively. The first panel highlights descriptive statistics, like the mean and standard deviations, while in the second one we report the results for three different tests. In the table,(ΔX_{t-1}) *corresponds to the change in the level of the target variable,* $e_{t-1}t$ *represents the forecast errrors and* $\Delta\hat{F}_t$ *the difference in the median forecast. The symbol* *** *stands for* $p < 0.01$.

the economy. In other words, forecasters appear to react to developments of the economy or changes in the accuracy of their forecasts by augmenting the disagreement surrounding the median forecast. This evidence is consistent with studies on other surveys (e.g., Andrade and Bihan 2013) therefore showing that Bloomberg and other types of professional forecasters behave similarly.

3.3.3 Surprises and some properties

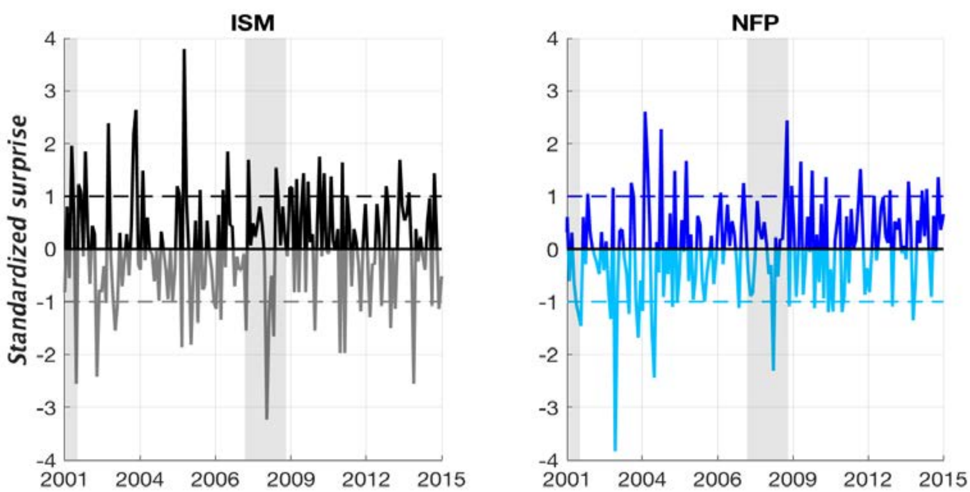
To create our surprise measure we follow the voluminous literature on macroeconomic releases and define the surprise as the standardized difference between the actual realization and the median expected value obtained by the panel of Bloomberg surveyed economists. More practically, we take the forecast errors described before but standardized by its standard deviation to make comparable the effect of different releases in a regression framework. Equation (3.2) shows the surprise measure we constructed

$$S_t = \frac{A_t - \hat{F}_t}{\sigma^s} \tag{3.2}$$

where A_t is the actual value of the release and $\hat{F}_t$ is its expected value obtained, as usually done in the literature, using the median of the distribution of the forecasters and σ^s is the standard deviation of the surprises over the sample in consideration.

Figure 3.3 presents the time series of the surprises over the sample in consideration. No particular pattern emerges from a simple eyeball analysis of the series. However, although

Fig. 3.3 ISM and NFP market-based surprises.



Note: for illustrative purpose, dark-color lines correspond to positive surprises while light colors to the negative ones. Dotted lines indicate a one-standard-deviation threshold.

predicting surprises per *sè* appears extremely complicated, we aim to investigate whether we can at least have some predictability for large surprises, i.e., greater than one standard deviation. The intuition is that some features of the forecasts distribution, such as its skewness, could give some insights into the size of the surprise. As shown by Félix et al. (2018), the skewness of the forecast distribution is strongly and positively linked to economic surprises, given that it could potentially be viewed as an indicator of extreme uncertainty regarding the release and therefore "anticipating" possible large surprises.

We formally test this hypothesis by proposing a simple probit model in which we use the skewness of the distribution at each release to predict the likelihood of a large surprise[6]. Analytically, this is equal to equation (3.3)

$$P(|s_t| > s) = \Phi(skewness(F_t) * \beta) \tag{3.3}$$

where s represents the size of the surprise which in this case spans from 1 to 1.5 (in absolute terms) standard deviations at interval of 0.1, and $\Phi(\cdot)$ is a standard cumulative normal distribution[7]. To evaluate the performance of this model, we first compute the receiver operating characteristic curve, also known as the ROC curve, which is created by plotting the true positive rate against the false positive rate at a various threshold level. Then, we use the area between the ROC curve and the naive 45-degree line (also known as the AUROC), to gauge the gain from using this model relative to a standard naive coin-flip strategy.

Figure 3.4 presents the ROC curves which are colored based on the size of the surprises with lighter colors indicating larger ones. A simple eyeball inspection shows that there is a small but non-negligible gain from using this simple model. In other words, the skewness of the distribution appears to have some explanatory power for large surprises for both variables, which is decreasing as the size of the surprises increases, especially in the case of NFP.

3.4 In-sample results

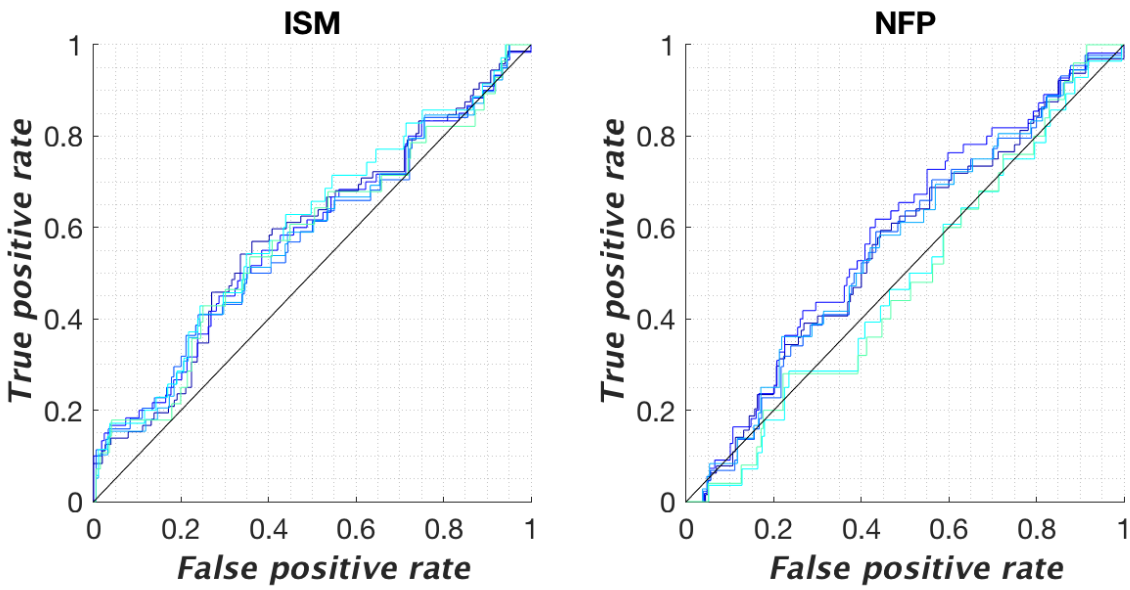
In this section we study the effects of these market-based surprises generated by NFP and ISM releases on the eurodollar spot rate. The framework we have in mind is the following: in a short window around the macroeconomic release (generally the first 60 minutes), movements in asset prices are only due to the unexpected component of the releases itself. Following this reasoning, we study a long time-series of releases in order to find patterns and exploit historical regularities to predict asset movements. To do so, we adopt an in-sample approach to test the main findings in the high-frequency literature. Loosely speaking, we show how to exploit the mixed-frequency nature of market-based surprises and asset movements to develop a framework to predict the asset price level.

Although in this paper we mainly focus on analyzing the effects of the US non-farm payroll and ISM market-based surprises on the eurodollar exchange rate, the framework we

[6] The exercise is implemented using an in-sample approach. The logic of this exercise is more oriented to identify possible drivers of large surprises rather than properly predicting them in an out-of-sample fashion, which would require to recalculate the standardized surprises after each release.

[7] The same exercise for surprises larger than 1.5 standard deviation becomes too rare to be tested.

Fig. 3.4 Receiver Operating Characteristic (ROC) curve – large surprises.



Note: each line represents the ROC curve computed for the model based on the different size of the standardized surprise. Lighter lines indicate the results for larger surprises.

present here is more general and suitable to any combination of macroeconomic releases and asset classes. Therefore, this framework can be taken as a method to guide market-operators trading on asset prices around macroeconomic releases.

3.4.1 The baseline model

Our investigation starts from a benchmark linear regression model which only includes a constant and the market-based surprise as independent variables, as shown in the following equation

$$y_t = \alpha + \beta S_t + \epsilon_t \tag{3.4}$$

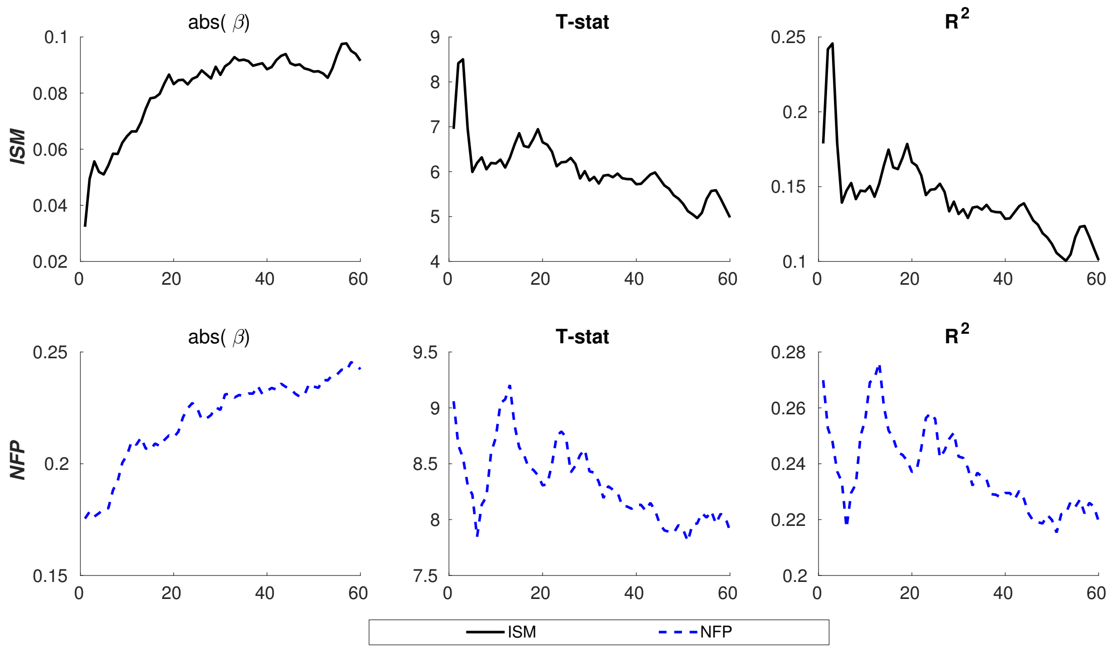
where y_t is the log-difference of the asset price in a window around the macroeconomic release, S_t is the standardised market-based surprise, as described in equation (3.2), and ϵ_t is a regression error with $\mathbb{E}_t(S_t, \epsilon_t) = 0$[8]. The dependent variable, i.e., the asset reaction, can be computed considering different starting point and different size of the window surrounding

[8]In case of stocks, exchange rates, and commodity prices it is common to report results as log-difference multiplied by 100 to reflect percentage changes while for assets bearing interest rates, like government bonds, are commonly reported in level-difference multiplied by 10000 to reflect movements in basis points.

the data release[9]. In our case, we compute the log-difference from the k minutes after the release, $k = 0, \ldots, 60$, to the minute before the release, therefore y_0 represents the reaction on impact and y_{60} the one-hour price change in the asset[10].

Figure 3.5 shows the results from equation (3.4), in which the first line refers to ISM (solid black line) while the second (blue dashed line) to NFP. The panels, starting from left, show the absolute values of the regression coefficients, the t-statistics, and the R^2. The first

Fig. 3.5 results from the linear regression model - $y_t = \alpha + \beta S_t + \epsilon_t$



Note: the figure shows the results from a linear regression model, as in equation (3.4), to a battery of dependent variables computed as a log-difference from the k minutes after the release to the minute before the release. The first line shows the results for the ISM (solid black lines), while the second for NFP (dashed blue lines). The panels, starting from left, show the absolute values of the regression coefficient, the t-statistics, and the R^2.

column presents some information regarding the magnitude of the asset movements after the surprise. As the surprises are standardized, the coefficient corresponds to the average change in the asset price at a specific k minute after a one standard deviation surprise in the corresponding release. Therefore, on average, it seems that the NFP release has a larger impact on exchange rates with respect to the ISM, consistently with the importance of this release also highlighted in the literature (Swanson and Williams, 2014). The second and

[9] In the literature there are many examples of specific windows employed to build the dependent variable; for example, Faust et al. (2007) employ a 5-15 minutes window, which is extended by Balduzzi et al. (2001) and Andersen et al. (2007) to 5-30 and 5-190 minutes, respectively.

[10] We restrict our focus to a 60-minute window because empirical works show that the effects of these surprises tend to disappear about 30-45 minutes after the release.

third columns show *via* the t-statistics and the R^2 that the estimated coefficients are always highly significant and these surprises explain a large fraction of asset movements.

Two interesting features are worth to highlight. First, the information content is stronger in the first minutes after the release. This is evident from the higher R^2 and t-statistics. Secondly, such a content starts rapidly decaying. This pattern is common for many macroeconomic releases and asset prices[11]. However, the main message of the chart is that, in order to create a dependent variable embedding the information content released in the macroeconomic announcement, a researcher should focus on a time-window no longer than one hour.

3.4.2 Non-linearities

Following the literature on macroeconomic news, this section shows different models exploiting specific characteristics considered crucial to determine asset price movements. In particular, we analyze: (i) the *asymmetry* in the effects of positive and negative surprises, (ii) the possible *non-linearity* of the effect of large surprises, (iii) the dependency on level of disagreement in the forecasters (which could be interpreted as a form of *state-dependency*), and (iv) the *time-variation* in the release effect (among others Andersen et al. 2007; Ehrmann and Fratzscher 2005; Faust et al. 2007; Galati and Ho 2003; Pericoli and Veronese 2015; Roache and Rossi 2010).

Now we describe our models' specification. First, equation (3.5) includes an indicator function accounting for differences in positive and negative shocks $\mathbb{I}_t^A$. This is equal to 1 when the surprise is positive and to -1 when the shock is negative; it is zero when the surprise is zero.

$$y_t = \alpha_1 + \beta_p S_t \mathbb{I}_t^A (S_t > 0) + \beta_n S_t \mathbb{I}_t^A (S_t < 0) + \epsilon_t \tag{3.5}$$

Secondly, equation (3.6) considers the effects generated by large surprises. In this case, *large* corresponds to the surprise larger than the 90^{th} percentile of the surprise distribution in absolute value. Thus, to account for these particular outcomes, we include an indicator function $\mathbb{I}_t^L$ which is 1 in case a surprise is large according to our definition, and it is zero in case of a standard surprise (smaller than the 90^{th} percentile in absolute value).

$$y_t = \alpha_2 + \beta_b S_t \mathbb{I}_t^L (S_t > 90^{th} pct(|S_t|)) + \beta_s S_t \mathbb{I}_t^L (S_t \leq 90^{th} pct(|S_t|)) + \epsilon_t \tag{3.6}$$

Third, following Pericoli and Veronese (2015), we extend the baseline model by accounting for state dependency in the form of disagreement among forecasters. The disagreement

[11] As an example, we report in the appendix the case of the stock market to corroborate this point. In particular, we show the Dow Jones stock index for the ISM, and the STOXX50E index for the NFP release. The reason for using a European index for the NFP release is because the US stock market is closed at the release time. This point reinforces why we are showing our application on the eurodollar exchange rate

is measured by estimating the cross-sectional standard deviation σ_t^D of the forecasters' distribution for each release A_t. Using the same approach, when $\sigma_t^D \in \mathbb{R}^+$ is larger than the $66^t h$ percentile of the forecasters' distribution, there is disagreement among the forecasters on the outcome of the release. The opposite is true for $\sigma_t^D \leq 66^{th} pct(\sigma_t^D)$. Accordingly, we construct an indicator function $\mathbb{I}_t^S$ which is equal to 1 when there is disagreement among forecasters, and zero otherwise. The two states are labeled *high* and *low* disagreement and are modelled as in equation (3.7).

$$y_t = \alpha_3 + \beta_h S_t \mathbb{I}_t^S(Disag_t > 66^{th} pct(\sigma_t^D)) + \beta_l S_t \mathbb{I}_t^S(Disag_t \leq 66^{th} pct(\sigma_t^D)) + \epsilon_t \tag{3.7}$$

Last but not least, we include an additional model to parsimoniously take into account the possible time variation of the coefficients in the model. We do so by including three time dummies to account for differences in the coefficients during the *pre-crisis period* from the beginning of the sample till 2007, the *crisis period* from 2007 till 2009, and finally the *quantitative easing* period from 2009 till the end of the sample, as done in (Pericoli and Veronese, 2015). This choice allows us to capture the possible time-variation in the effects but remaining parsimonious in the number of coefficients (for example Swanson and Williams, 2014). The model is shown in equation (3.8)

$$y_t = \alpha + \sum_j \delta_j \beta D_t^j S_t + \epsilon_t \tag{3.8}$$

where D_t^j with $j \equiv T, C, Q$ is a set of dummy variable capturing the pre-crisis period from the beginning of the sample till 2007, the crisis period from 2007 to 2009 and the post-crisis or quantitative easing period from 2009 till the end of the sample. δ_j is the coefficient linked to the dummy variables that multiplicative enter into the equation and allows $\delta_j \beta$ to change in different periods.

Table 3.3 shows the regression coefficients and the t-statistics for the models described in equations (3.4) to (3.8). The coefficients in the tables can be read as follows; a one standard deviation NFP surprise increase/decrease the eurodollar exchange rate by x basis points. For example, the linear model suggests that a NPF surprise lowers the exchange rate by 17 basis points. As the main interest in the paper is forecasting, we do not dig deeper into the results regarding the magnitude of the macroeconomic surprises on asset prices. However, consistently with the findings in the high-frequency literature, we find all the slope coefficients statistically significant, implying that around the releases, the variation in the asset price is mostly explained by the market-based surprise, while the estimated constants are always not significantly different from zero. Secondly, the estimated effects on the eurodollar are always negative. The reason is that a positive (negative) surprise in both ISM and NFP is considered as a good (bad) news for the US economy, triggering an inflow of capitals which

Table 3.3 ISM and NFP regression coefficients.

	EUR^{ISM}	t-stat	EUR^{NFP}	t-stat
1. Linear Model, $y_t = \alpha + \beta S_t + \epsilon_t$				
α_0	0.02	0.04	-2.57	1.35
β	-3.24	6.95	-17.55	9.06
2. Asymmetric model, $y_t = \alpha_1 + \beta_p S_t \mathbb{I}_t^A(S_t > 0) + \beta_n S_t \mathbb{I}_t^A(S_t < 0) + \epsilon_t$				
α_1	-0.50	0.69	1.71	0.59
β_n	-2.61	3.22	-25.39	5.71
β_p	-3.98	4.33	-13.04	4.34
3. Size model, $y_t = \alpha_2 + \beta_b S_t \mathbb{I}_t^L(S_t > 90^{th}pct(\lvert S_t \rvert)) + \beta_s S_t \mathbb{I}_t^L(S_t \leq 90^{th}pct(\lvert S_t \rvert)) + \epsilon_t$				
α_2	0.05	0.10	-2.17	1.14
β_b	-2.57	4.43	-15.44	7.35
β_s	-4.39	5.75	-27.89	5.99
4. State-dependent model, $y_t = \alpha_3 + \beta_h S_t \mathbb{I}_t^S(Disag_t > 66^{th}pct(\sigma_t^D)) + \beta_l S_t \mathbb{I}_t^S(Disag_t \leq 66^{th}pct(\sigma_t^D)) + \epsilon_t$				
α_3	0.03	0.06	-2.58	1.34
β_h	-1.80	2.53	-17.67	5.64
β_l	-4.26	7.10	-17.47	7.18
5. Time-variation model, $y_t = \alpha + \sum_j \delta_j \beta D_t^j S_t + \epsilon_t$				
α_4	0.04	0.09	-2.32	1.20
β_1	-2.92	4.61	-17.02	7.13
β_2	-3.49	2.52	-11.83	1.53
β_3	-3.66	4.59	-19.89	5.51

Note: the table shows the values and t-statistics for the parameters estimated in equations (3.4) *to* (3.8). *The coefficients are reported in basis points–i.e., log-change times 10000.*

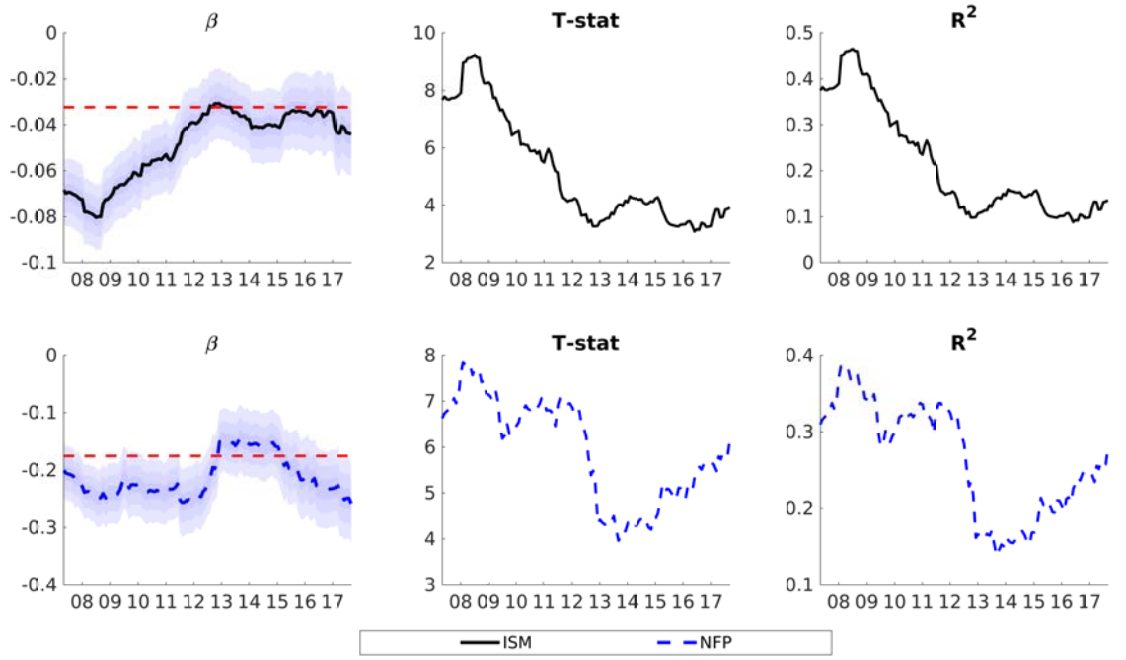
makes the dollar more competitive, meaning that less (more) dollar are required to purchase euros. Last, there is evidence that the various forms of non-linearities have different effects on the asset price, and therefore we use this finding with our out-of-sample forecast analysis in the next section.

3.4.3 Rolling-window estimation

Including time-variation in the market response can provide a better gauge of the strength of the surprise along different periods. Practically, this part extends the model (3.8) by allowing for rolling-window estimation of the linear regression model.

Figure 3.6 shows the variation of the first-minute coefficients estimated according to this methodology for the benchmark model specified in equation (3.2). We select a starting sample of $T = 100$ observations, then we remove one observation at the very beginning of the sample and add one at the very end, and repeat the procedure estimating and saving the first-minute coefficients for the entire sample. The first row of the figure shows the results for ISM while the second for NFP. The first column shows the evolution of the coefficients along time supported with confidence bands (95, 84, 66 percentiles from the distribution of β). The horizontal dashed red lines report the coefficients estimated using the full sample.

Fig. 3.6 rolling-window regression.



Note: the figure shows the results from a rolling-window regression model, as in equation (3.4)*. The dependent variable is constructed selecting the minute which maximizes the R^2 in a full-sample regression. The first line shows the results for the ISM (solid black lines), while the second for NFP (dashed blue lines). The panels, starting from left, show the absolute values of the regression coefficient, the t-statistics, and the R^2.*

The second column shows the t-statistics for the estimated betas and the last column the corresponding R^2.

The results confirm the findings in the full sample regression, and in particular the fact that the market-based surprise is a crucial determinant of the asset price around a short window. Also, figure 3.6 displays other essential characteristics contingent on the two variables we are analyzing. In particular, both releases exert significant time variation in the effects of the euro-dollar exchange rate. However, although the NFP effect is more stable around the estimated full sample coefficient, the ISM surprise presents an increasing at the beginning of the sample, then stabilize.

3.5 Out-of-sample results

This section presents an exercise to test whether the *in*-sample fit found in the previous section and extensively studied in the literature holds switching to a real-time *out*-of-sample approach. To do so, we develop a framework to assess the performance of the models described by equations (3.4) to (3.8). We show that although non-linearities and time-variation seem to be essential features for the asset price responses to a market-based surprise, these characteristics do not add any strong and robust predictive power to regression models in a *out*-of-sample framework. Consequently, we propose some possible reasons for this results and we study whether there could be some components which can be systematically predicted.

Overall, our contribution in this section is twofold: on the one hand, we show and compare the performance of the different models highlighted in the in-sample literature within an out-of-sample setting. On the other, we develop a framework to assess and help market-operators to analyze in real-time the likely behavior of an asset in a short window after a market-based surprise.

3.5.1 Building an out-of-sample framework for market-based surprises

The frequency of the market-based surprises can be weekly, monthly or quarterly depending on the corresponding indicator. However, asset prices are traded in real-time and quotes are available at tick level. Therefore, although the macroeconomic release is considered contemporaneous to the asset movements at the frequency of the market-based surprise, exploiting the tick structure of the asset price data, it is possible to forecast the direction of the asset price in a short window after the release, conditional to the release itself.

An example can help to clarify: suppose that a market operator is interested in predicting the effect of the NFP surprise on the eurodollar, as in our case. She collects all the historical data about NFP surprises and changes in the eurodollar the minute before and the minute of the release and trains a model to estimate the effect of such a surprise[12]. However, the problem is that as soon as the new NFP surprise is known, also the dependent variable is known and there is no longer need for prediction.

There are at least two ways to deal with the issue of contemporaneity in this particular setting. First, the market operator can attempt to predict the dependent variable without updating the model as shown in equations (3.9) and (3.10) where data are available up to $T-1$ and the next macroeconomic release is at time T.

$$y_{1:T-1} = \beta^{(T-1)} S_{1:T-1} + \epsilon_{1:T-1} \tag{3.9}$$

$$\hat{y}_T = \beta^{(T-1)} S_T \tag{3.10}$$

The notation $y_{1:T-1}$ refers to a $(T-1)\times 1$ vector of one minute asset price changes, $S_{1:T-1}$ is a $(T-1)\times 1$ vector of standardized market-based surprises, $\epsilon_{1:T-1}$ is a $(T-1)\times 1$ vector

[12]The price attached to a particular minute is usually the closing price within the minute. Therefore, the effect of the surprise is embedded in the same minute in which the release is released.

of regression errors, and the subscript $1:(T-1)$ stands for observations from time $t=1$ to $t=T-1$. Finally, $\beta^{(T-1)}$ is the slope estimated using this particular sample size. $\hat{y}_T$ is the predicted value for the asset price at minute frequency. This corresponds to combine the market-based surprise at time T with the coefficient estimated using data up to $T-1$, $\beta_{1:T-1}$. However, this procedure can be hardly implementable in real-time because it requires to update the model, predicting $\hat{y}_T$ and perform a transaction in a fraction of second. A more natural way would be to simulate an $(S \times 1)$ vector $S_T^{(s)}$, where s is the number of simulated surprise magnitude, $s=1,\ldots,S$, of plausible values of the surprise S_T and derive some summary features of the implied asset price movement as minimum, maximum and quantiles. This method has the clear advantage of providing guidance much in advance than the actual market-based release[13].

However, based on the first methodology, a second one is directly available by assuming that the effect of the market-based surprise on the asset price presents some persistence in a short period after the release. In our study, such a period is considered equal to sixty minutes. This assumption is plausible, as shown in figure 3.5, after this period the regression coefficient is still significant and the R^2 relatively high. Therefore, we can extend the methodology displayed in equation (3.9) and (3.10) by using a system of linear regressions as shown in equation (3.11) and (3.12)

$$y_{1:T-1}^{\tau-1:\tau+k} = \beta_{\tau-1:\tau+k}^{(T-1)} S_{1:T-1} + \epsilon_{1:T-1}^{\tau-1:\tau+k} \tag{3.11}$$

$$\hat{y}_T^{\tau-1:\tau+k} = \beta_{\tau-1:\tau+k}^{(T-1)} S_T \tag{3.12}$$

where $y_{1:T-1}^{\tau-1:\tau+k}$ is a $(T-1)\times K$ array of K minutes asset price changes, $\epsilon_{1:T-1}^{\tau-1:\tau+k}$ is a $(T-1)\times K$ array of regression errors, and the superscript ${}^{\tau-1:\tau+k}$ stands for changes betwen the minutes before the revelation of market-based surprise and $k=0,\ldots,K$ minutes after. $\beta_{\tau-1:\tau+k}^{(T-1)}$ is a vector of $(K\times 1)$ regression coefficients. Finally, $\hat{y}_T^{\tau-1:\tau+k}$ is a $(K\times 1)$ vector of predictions. According to this methodology, the objective of our prediction is $y_T^{\tau-1:\tau+k}$, which can be considered as a cumulated impulse response function due to the market-based surprise.

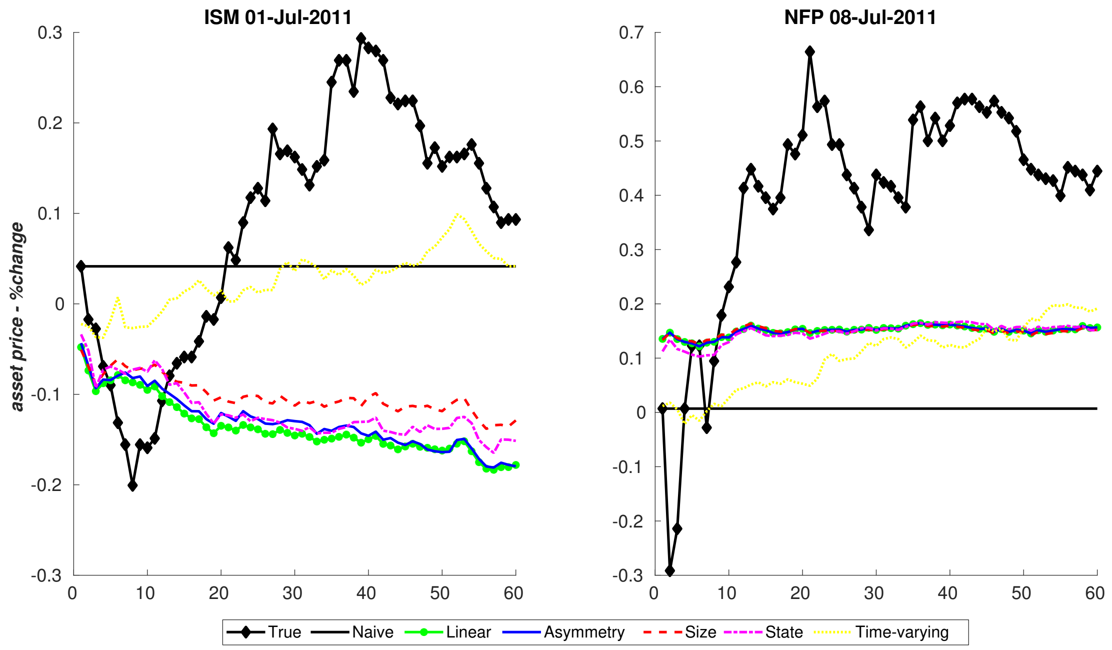
3.5.2 Quantitative results

To assess the predicting ability of the different models described in equation (3.4) to (3.8), we perform the following experiment: (i) we estimate each model on a sample starting from February-1999 up to September-2010, and (ii) we pre-estimate the coefficients $\beta_{\tau-1:\tau+k}^{(T-1)}$; (iii)

[13]The main advantage of this methodology is that a market agent can set up some simple algorithms based on *if* conditions. Thus, as soon as the real surprise comes out, the algorithm can immediately put in place a transaction without any need to update the model or computing the prediction.

we recursively estimate the $\hat{y}_T^{\tau-1:\tau+k}$ in a one-step ahead out-of-sample fashon for a ten years period–i.e., from September-2010 to September-2017.

Fig. 3.7 actual data and predicted changes for two selected ISM and NFP dates.



Note: the figure shows the predicted path for the first thirty minutes after the release made by the models described in equations (3.4) to (3.8), plus a constant model (green horizontal line) for two selected dates of the ISM (left panel) and NFP (right panel) release. The constant model is fixed at the level of the first-minute change after the release. The dashed black lines display real data.

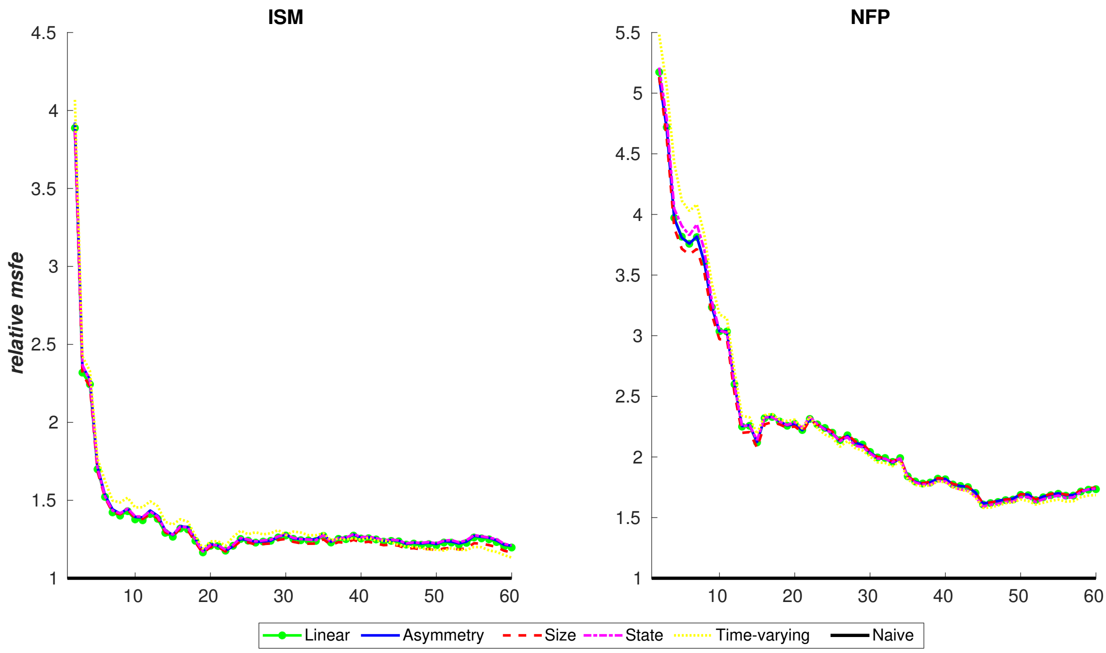
The process implies that for each release date, we estimate an entire 60-minute path and not a single point. To understand this point, figure 3.7 shows a selection of two dates for the predicted path of the eurodollar exchange rate after the ISM and NFP releases. To evaluate the goodness of the different prediction, we compare the estimated models against a "naive" model which is built by holding constant the change in the asset price at the level of the minute of the release (dashed green line). This constant model is the simplest baseline model one can think of, and, in market jargon, it would be equivalent to betting on a short-term level shift at the level of the initial change. When the asset price remains in a neighborhood of the post-release value, the constant model will have a tiny error. On the opposite, when the asset price goes back to the starting level (mean reversion), or move in the opposite direction, the error would be significant. The model is evaluated according to the mean squared forecast error in the form of equation (3.13) over the 84 release dates.

$$MSFE_{\tau-1}^{\tau+k} = \frac{1}{T}\sum_{t=T0}^{T}\left(y_t^{\tau-1:\tau+k} - \hat{y}_t^{\tau-1:\tau+k}\right)^2, \quad k = 0,\ldots,K \tag{3.13}$$

where $MSFE_{\tau-1}^{\tau+k}$ is a $K \times 1$ vector meaning that we evaluate each k minute after the release.

Figure 3.8 graphically shows the relative $MSFE_{\tau-1}^{\tau+k}$ of the considered set of models against the naive model (horizontal dashed green line at 1) for the first sixty minutes after the release. For each minute, when the relative MSFE is larger than one, the naive model is more informative than the considered model. When the relative MSFE is less than one, the opposite is true. The left-hand panel shows the case of the ISM, while the right-hand one the NFP. The central message from the two panels is clear: the naive model has a performance

Fig. 3.8 relative mean squared forecast error

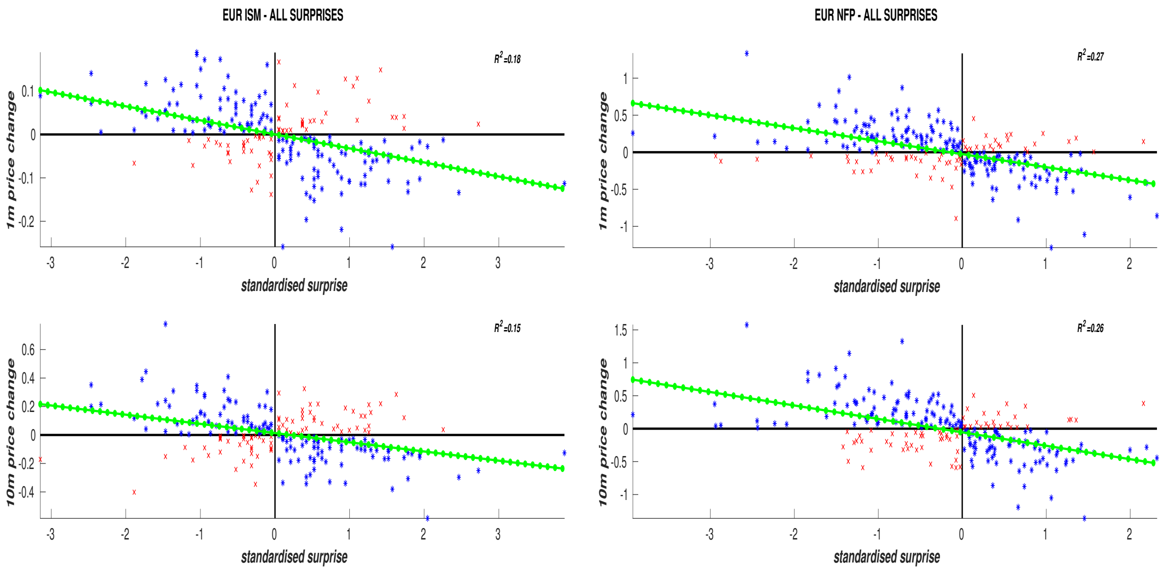


Note: the figure shows the relative mean squared forecast error of the predictions made by the models described in equations (3.4) to (3.8) against a constant model (green horizontal line). When the relative MSFE is above the green horizontal line, the constant model performs better.

not comparable with any of the considered models. Secondly, we notice that the distance between the naive and the other models shrinks rapidly. This feature is related to the fact that at minute zero, the error of the naive model is equal to zero by construction. Therefore, in the first few minutes after the release, unless the asset price keeps moving widely, the naive model would have a considerable advantage. However, still, after sixty minutes, our set of models, on average, are outperformed by the naive model.

These results shed light on the difficulties that these models have in detecting the reaction on impact in the asset price which also affects the response over the following minutes. To understand why this is the case, we look more in details at the data. In particular, we point out that in some periods the change in the asset price goes against what implied by economic theory, i.e., positive surprises lead the dollar to depreciate and viceversa. Figure 3.9 shows the scatter plots of the standardized ISM and NFP surprises against the asset price changes in one (top panels) and ten (bottom panels) minutes after the macroeconomic release together with a regression line. In theory, a positive surprise in both the ISM and NFP releases

Fig. 3.9 ISM and NFP scatter plots



Note: the figure shows the scatter plot for the market-surprise and the eurodollar variation. In all four panels, the blue stars display the points in which the asset price behaves as expected, while the red crosses when it has a counterintuitive movement. The first column highlights eurodollar variation after 1 and 10 minutes after the release of the ISM, while the second column after the NFP.

should cause a reaction in the eurodollar of the opposite sign, because a positive surprise in these indicators is considered as a good news for the US economy, and as a consequence it should lead to an appreciation of the USD (less USD to buy one euro). Therefore, the scatter plots should present points only in the second and fourth quadrant, which would represent a response in the "correct direction" (blue stars). However, the charts show that in many occasions the market reaction has the same direction as the asset price change implying a response in the "wrong direction" (highlighted by red crosses). This reaction has a substantial effect on the coefficient estimated in our models, as the slope will rotate, and the effect underestimated.

This enlighting consideration helps explaining why out-of-sample models are not particularly performing. The underline reason is that, after a macroeconomic release, even the direction of the asset price, which should be straightforward, is not easy to predict. Consequently, as predicting the magnitude of the asset price is a more challenging task, the fact that the prediction is poor comes as a direct consequence, especially in an out-of-sample framework. Nevertheless, the evidence that the most considerable part of the market response is showed to be on the right side induces us to run a second experiment, to see whether we can identify some common characteristics relating the "correct-responses".

3.5.3 Cluster analysis – it's in the size

In what follows, we investigate the possibility of having at least some predictability for certain *type* of surprises. Specifically, we expect that the direction of the price change might go against the theory when the surprise is small while it should always be correct for large surprises. To test this hypothesis, we cluster the "population" of the asset impulse response functions according to the magnitude of the standardized market-based surprise, calculating the standard deviation using the whole sample.

In particular, we create ten different contiguous bins and assign each asset response to a single one. The bins range from minus to plus two standard deviations (σ^s), with a step of 0.5. For each cluster, we assess both the average direction and magnitude. As each member of the population is a $T \times 1$ vector containing the asset price response to a particular surprise, the entire population for each bin will be a matrix $T \times K^b$, where K^b is the number of market-based surprises populating the bin $b = 1, \dots, 10$. By construction, bins corresponding to clusters of large standard deviations will be less populated than the others.

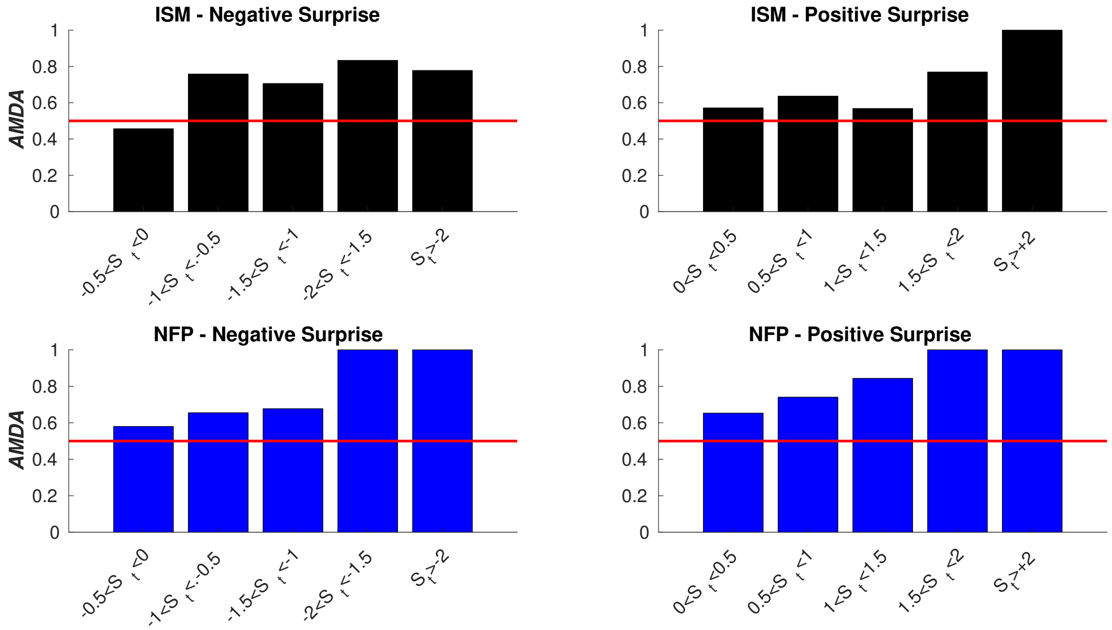
To highlight some characteristics of the clusters, we construct a measure of average mean directional accuracy (AMDA), which is described by equation (3.14):

$$AMDA^b = \frac{1}{KT^b} \sum_{k=1}^{K} \sum_{t=1}^{T^b} \Big(sign(y_t^{\tau-1:\tau+k}) \neq sign(S_t) \Big), \quad b = 1, \dots, 10 \,\text{bins} \tag{3.14}$$

where T^b denotes the number of surprises in the bin b. In particular, this measure shows the frequency with which the asset price moves on average in the "correct" direction in the first sixty minutes after the release[14]. The AMDA is reported in figure 3.10 for the ISM (first row, black bars) and NFP (second row, blue bars). We present four panels to divide the response to positive and negative surprises, presented according to our ten bins. More specifically, for each cluster, we compute the AMDA, and each bar represents the average direction of the sample of prices in the first sixty minutes after the surprise. To assess our results, we use a *naive* guess represented by the red horizontal line corresponding to an $AMDA = 0.5$. The chart confirms the idea that the responses to large surprises tend to be of the "correct" direction while. Indeed, the bars are generally well above the red line for large surprises, and this is common for both sign and releases.

[14] Remember that the signs in equation (3.14) has to be different because positive (negative) surprises are associated with an appreciation (depreciation) of the dollar with respect to the euro.

Fig. 3.10 average sign predictability



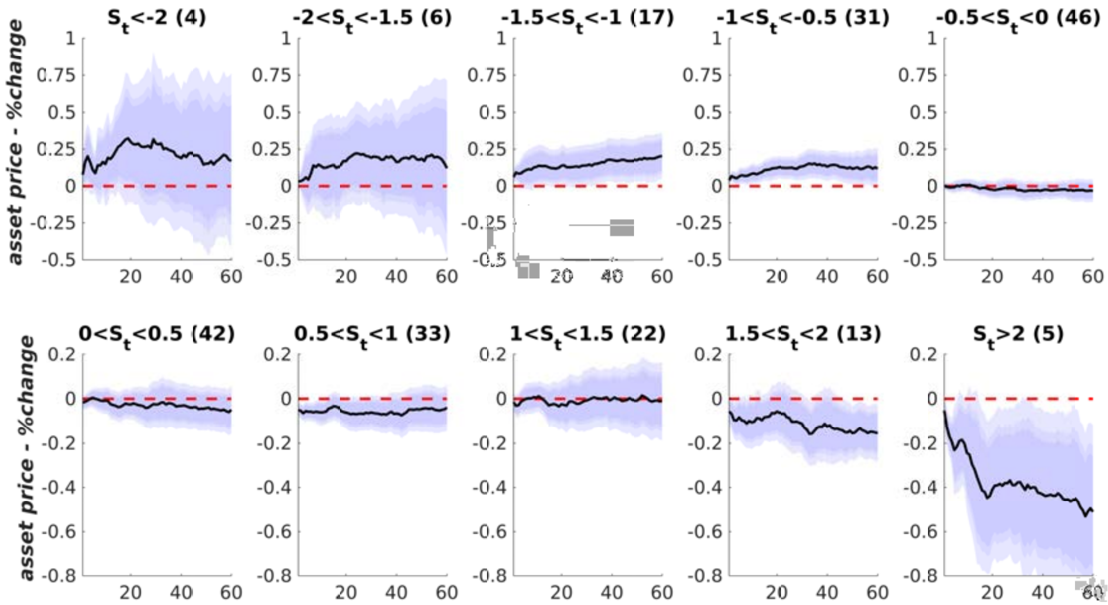
Note: each panel of the figure shows the frequency with which the asset price moves on average in the "correct" direction in the first sixty minutes after the ISM (first row, black bars) and NFP (second row, blue bars) releases. The red horizontal line highlight $AMDA = 0.5$. At this level, on average, the asset direction can be considered as equivalent to a random guess.

Moved by this finding, we use the same approach to analyze the size of the asset price responses and to see whether they follow the same pattern. Theoretically, they should, as they are driven by the average direction of the impulses. Besides, this helps to test whether the average response is statistically different from zero, which is what a market agent should be interested in. Suppose there is a consistent pattern for a response in a certain bin, for example negative surprise larger than two standard deviations always lead to a statistically significant average increase in the asset price in the first sixty minutes after the release, then market agents can exploit this pattern to secure some profits.

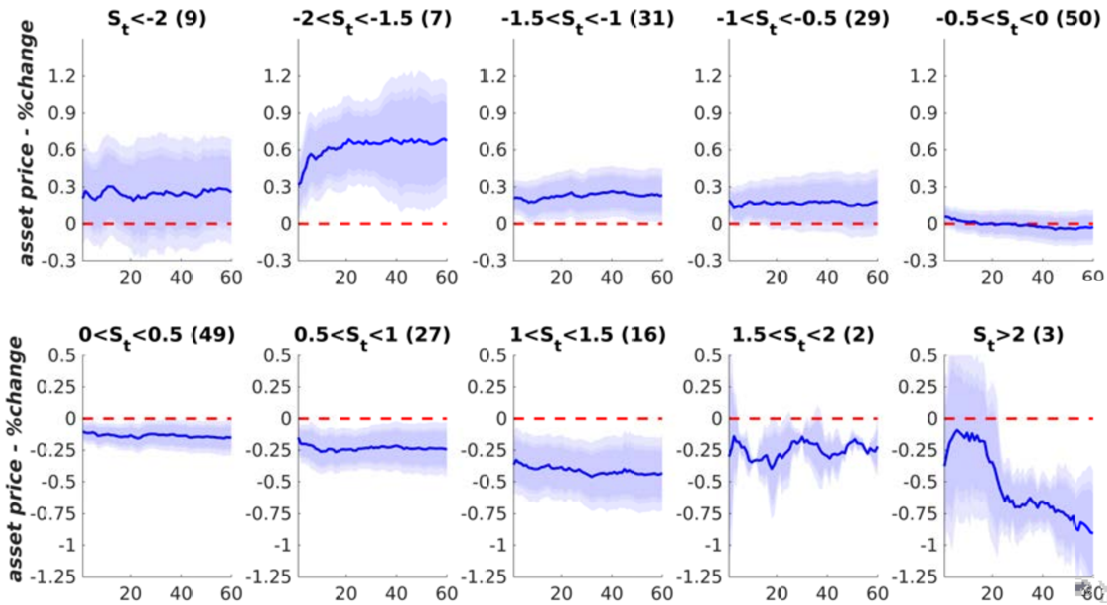
Figure 3.11a and 3.11b show the median response of the asset price population in each of the described bins (solid blue line) and the relative confidence bands for the 1, 5 and 10 percent significance levels (shaded areas). As for figure 3.10, the chart shows in each panel one of the ten bins in which the surprise response is clustered. These charts confirm the tendency to move toward a more clear pattern as the size of the surprise increases. In addition to the previous charts, the confidence bands provide evidence on whether the response is not different from zero (which can be thought as the "random-guess" in the previous case).

However, there is an important caveat regarding large surprises, because a small number of episodes populates those bins, so the responses might not be precisely estimated, and

(a) ISM - median response



(b) NFP - median response



Note: the figure shows the average response (solid blue lines) of the asset price in the first sixty minutes after the market-based surprise clustered in 10 different bins. It also shows the 99, 95 *and* 90*th confidence bands as shaded areas.*

market operators should include some judgment in interpreting the result coming from this procedure. Nevertheless, having the possibility to know in real-time the history of the asset population concerning a specific market-based surprise dimension as soon as the release is revealed provides a valuable framework to help taking investment decisions.

In our experience, many macroeconomic releases present a pattern which is remarkably similar to the one presented using the ISM and NFP surprises. In fact, on the one hand, the framework provides a data-driven help in also assessing asset price variations due to less well-known releases and vice-versa. On the other hand, it allows for a "search for market inefficiencies" which can be exploited by market operatotors.

3.6 Conclusions

This paper proposes a forecasting exercise that focuses on the possible prediction of the effects of macroeconomic surprises. We empirically investigate whether the possible predictability obtained using an in-sample approach claimed by a vast literature (Andersen et al., 2007; Ehrmann and Fratzscher, 2005; Fair, 2003; Pericoli and Veronese, 2015) also holds within a real-time out-of-sample exercise. To do so, we use as a laboratory two of the most important macroeconomic releases that impact the US financial markets, as the change in the level of non-farm payrolls and the ISM Manufacturing survey, and we explore the reaction of the euro-dollar exchange rate using minute-by-minute data.

We summarize our findings in a few points. First, non-linearities and time-variation matter because they have some predictive power when used in an in-sample approach. Secondly, the in-sample predictability deteriorates when we switch to a real-time out-of-sample exercise, and we show that a constant model is hardly outperformed. We also show that the reason is linked to the number of market-responses in the "correct" and "wrong" direction. This feature makes harder even to forecast the sign of the market reaction. Third, we highlight the fact that there is some predictability in the size and sign of market responses but only in a few numbers of cases. In particular, we cluster the surprises in different bins according to the size, and we show that larger surprises are easy to predict. This finding provides some linkage to relate the in-sample importance of nonlinearities consistently highlighted in the literature, to our out-of-sample results.

Finally, as the main contribution, this paper provides a framework that can be exploited by market agents to study the behavior of different asset classes around any macroeconomic release. In our experience, many pairs of assets and macroeconomic releases behave similarly, and present predictability which improves as the surprise size increases. However, we do not exclude that for less popular asset classes and releases, the predictability can be higher. In that case, besides providing a structure to guide market agents in real-time, our framework could also be used to set-up automatic trades based on the surprise/asset reaction history.

Appendix

Evidence on Seasonality

We estimate a monthly-dummy regression to test for the hypothesis of a possible pattern in some months (generally related to seasonal adjustments). May and August turn out to have a 10% statistically significant negative bias.

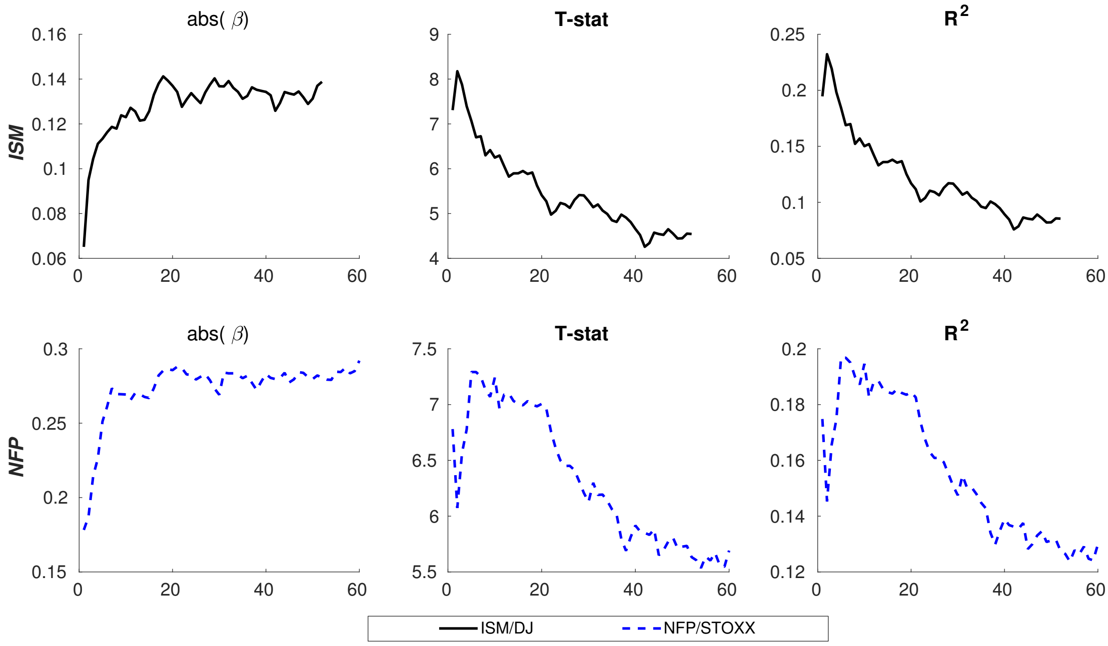
Table 3.1 monthly seasonality - $e_t = \sum_{i=1}^{12} \alpha_i D_{it} + u_t$

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
NFP												
α_i	-0.44	-8.84	-28.28	10.52	-35.84**	-27.42	-12.31	-37.89**	-29.75	14.63	-20.88	-21.72
t-stat	0.02	-0.48	-1.56	0.58	-1.98	-1.51	-0.68	-2.09	-1.60	0.78	-1.12	-1.17
ISM												
α_i	0.34	0.54	0.11	0.04	-0.04	0.48	-0.16	0.25	0.21	0.01	0.07	0.06
t-stat	0.77	1.22	0.26	0.10	-0.10	1.07	-0.37	0.56	0.47	0.01	0.16	0.15

** $p < 0.05$

Additional figures

Fig. 3.1 Dow Jones and STOXX50E. Results from the linear regression model.



Note: the figure shows the results from a linear regression model, as in equation (3.4), to a battery of dependent variables computed as a log-difference from the k minutes after the release to the minute before the release. The first line shows the results for the ISM on the Dow Jones (solid black lines), while the second for NFP on the STOXX50E (dashed blue lines). The panels, starting from left, show the absolute values of the regression coefficient, the t-statistics, and the R^2.

References

Aastveit, K. A., Gerdrup, K. R., Jore, A. S., and Thorsrud, L. A. (2014). Nowcasting GDP in Real Time: A Density Combination Approach. *Journal of Business & Economic Statistics*, 32(1):48–68.

Abel, J., Rich, R., Song, J., and Tracy, J. (2016). The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters. *Journal of Applied Econometrics*, 31(3):533–550.

Adrian, T., Boyarchenko, N., and Giannone, D. (2016). Vulnerable growth. Technical Report 794, Federal Reserve Bank of New York.

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2010). New Eurocoin: Tracking Economic Growth in Real Time. *The Review of Economics and Statistics*, 92(4):1024–1034.

Alvarez, R., Camacho, M., and Perez-Quiros, G. (2016). Aggregate versus disaggregate information in dynamic factor models. *International Journal of Forecasting*, 32(3):680–694.

Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, 25:177–190.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Vega, C. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, 73(2):251–277.

Andrade, P. and Bihan, H. L. (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, 60(8):967 – 982.

Antipa, P., Barhoumi, K., Brunhes-Lesage, V., and Darné, O. (2012). Nowcasting German GDP: A comparison of bridge and factor models. *Journal of Policy Modeling*, 34(6):864–878.

Aprigliano, V., Ardizzi, G., and Monteforte, L. (2017). Using the payment system data to forecast the Italian GDP. Temi di discussione (Economic working papers) 1098, Bank of Italy, Economic Research and International Relations Area.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Quarterly Review*, (Win):2–11.

Balduzzi, P., Elton, E. J., and Green, T. C. (2001). Economic news and bond prices: Evidence from the us treasury market. *Journal of financial and Quantitative analysis*, 36(4):523–543.

Barhoumi, K., Darné, O., and Ferrara, L. (2009). Are disaggregate data useful for factor analysis in forecasting French GDP? Technical report.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. *Handbook of Economic Forecasting, Elsevier.*

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1).

Brave, S., Butters, R. A., and Justiniano, A. (2016). Forecasting Economic Activity with Mixed Frequency Bayesian VARs. Working Paper Series WP-2016-5, Federal Reserve Bank of Chicago.

Busetti, F., Caivano, M., and Rodano, L. (2015). On the conditional distribution of euro area inflation forecast. *Temi di discussione (Economic working papers) 1027, Bank of Italy.*

Campbell, S. D. and Sharpe, S. A. (2009). Anchoring Bias in Consensus Forecasts and Its Effect on Market Prices. *Journal of Financial and Quantitative Analysis*, 44(02):369–390.

Carriero, A., Clark, T. E., and Marcellino, M. (2015). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society Series A*, 178(4):837–862.

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Ciccarelli, M. and Osbat, C. (2017). Low inflation in the euro area: Causes and consequences. (181).

Coibion, O. and Gorodnichenko, Y. (2012). What Can Survey Forecasts Tell Us about Information Rigidities? *Journal of Political Economy*, 120(1):116–159.

D'Agostino, A., Giannone, D., and Gambetti, L. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82–101.

D'Agostino, A., McQuinn, K., and O'Brien, D. (2012). Now-casting irish gdp. *Journal of Business Cycle Measurement and Analysis*, 2:21–31.

Dahlhaus, T., Guénette, J.-D., and Vasishtha, G. (2017). Nowcasting BRIC+M in real time. *International Journal of Forecasting*, 33(4):915–935.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–883.

Dovern, J. (2013). When are GDP forecasts updated? Evidence from a large international panel. *Economics Letters*, 120(3):521–524.

Ehrmann, M. and Fratzscher, M. (2005). Exchange rates and fundamentals: new evidence from real-time data. *Journal of International Money and Finance*, 24(2):317–341.

Elliott, G. and Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, 46(1):3–56.

Fair, R. C. (2003). Shock effects on stocks, bonds, and exchange rates. *Journal of International Money and Finance*, 22(3):307–341.

Faust, J., Rogers, J. H., Wang, S.-Y. B., and Wright, J. H. (2007). The high-frequency response of exchange rates and interest rates to macroeconomic announcements. *Journal of Monetary Economics*, 54(4):1051–1068.

Félix, L., Kräussl, R., and Stork, P. (2018). Predictable biases in macroeconomic forecasts and their impact across asset classes. *Mimeo*.

Galati, G. and Ho, C. (2003). Macroeconomic news and the euro/dollar exchange rate. *Economic notes*, 32(3):371–398.

Garcia, J. A. (2003). An introduction to the ECB?s survey of professional forecasters. Occasional Paper Series 08, European Central Bank.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55:665–676.

Greene, W. H. (2008). *Econometric Analysis*. International Student Edition, Fifth Edition.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American Economic Review*, 95(1):425–436.

Jarocinski, M. and Bobeica, E. (2017). Missing disinflation and missing inflation: the puzzles that aren't. (2000).

Kenny, G., Kostka, T., and Masera, F. (2014). How Informative are the Subjective Density Forecasts of Macroeconomists? *Journal of Forecasting*, 33(3):163–185.

Kenny, G., Kostka, T., and Masera, F. (2015). Can Macroeconomists Forecast Risk? Event-Based Evidence from the Euro-Area SPF. *International Journal of Central Banking*, 11(4):1–46.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

Li, W., Wong, M. C., and Cenev, J. (2015). High frequency analysis of macro news releases on the foreign exchange market: A survey of literature. *Big Data Research*, 2(1):33 – 48. Special Issue on Computation, Business, and Health Science.

Mankiw, N. G. and Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Manzan, S. and Zerom, D. (2013). Are macroeconomic variables useful for forecasting the distribution of u.s. inflation? *International Journal of Forecasting*, 29(3):469–478.

Marcellino, M., Porqueddu, M., and Venditti, F. (2016). Short-Term GDP Forecasting With a Mixed-Frequency Dynamic Factor Model With Stochastic Volatility. *Journal of Business & Economic Statistics*, 34(1):118–127.

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18:427–443.

Ottaviani, M. and Sorensen, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics*, 81(2):441–466.

Pericoli, M. and Veronese, G. (2015). Forecaster heterogeneity, surprises and financial markets. *Bank of Italy, Working Paper 1020*.

Pettenuzzo, D. and Timmermann, A. (2017). Forecasting macroeconomic variables under model instability. *Journal of Business & Economic Statistics*, 35:183–201.

Poncela, P. and Ruiz, E. (2016). Small- Versus Big-Data Factor Extraction in Dynamic Factor Models: An Empirical Assessment. In *Dynamic Factor Models*, volume 35 of *Advances in Econometrics*, pages 401–434. Emerald Publishing Ltd.

Roache, S. K. and Rossi, M. (2010). The effects of economic news on commodity prices. *The Quarterly Review of Economics and Finance*, 50(3):377–385.

Rossi, B. and Sekhposyan, T. (2015). Alternative Tests for Correct Specification of Conditional Predictive Densities. (758).

Schorfheide, F. and Song, D. (2015). Real-Time Forecasting With a Mixed-Frequency VAR. *Journal of Business & Economic Statistics*, 33(3):366–380.

Scotti, C. (2016). Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises. *Journal of Monetary Economics*, 82(C):1–19.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Stock, J. H. and Watson, M. W. (2010). Modeling inflation after the crisis. *NBER Working Paper 16488*.

Swanson, E. T. and Williams, J. C. (2014). Measuring the effect of the zero lower bound on medium-and longer-term interest rates. *American Economic Review*, 104(10):3154–85.

Tagliabracci, A. (2018). All against the median: testing the accuracy of the bloomberg consensus. *Mimeo*.

Tillmann, P. and Wolters, M. H. (2015). The changing dynamics of us inflation persistence: a quantile regression approach. *Studies in Nonlinear Dynamics & Econometrics*, 19(2):161–182.

Tsong, C.-C. and Lee, C.-F. (2011). Asymmetric inflation dynamics: Evidence from quantile regression analysis. *Journal of Macroeconomics*, 33:668–680.