

Central bank macroeconomic forecasting during the global financial crisis: the European Central Bank and Federal Reserve Bank of New York Experiences*

Lucia Alessi[†] Eric Ghysels[‡] Luca Onorante[§] Richard Peach[¶] Simon Potter^{||}

January 31, 2014

Abstract

This paper documents macroeconomic forecasting during the global financial crisis by two of the key central banks: the Federal Reserve Bank of New York and the European Central Bank (hereafter respectively FRBNY and ECB). The paper is the result of a collaborative effort between the two institutions, allowing us to study the time-stamped forecasts as they were made throughout the crisis. The analysis does not exclusively focus on point forecast performance. It also examines density forecasts provided by the FRBNY, the relative performance of various models used by the ECB as well as methodological contributions. It also provides insights on how financial market data can be incorporated into the forecasting process.

*The second author gratefully acknowledges the financial support from the European Central Bank, Federal Reserve Bank of New York and a Marie Curie FP7-PEOPLE-2010-IIF grant. Special thanks to M. Henry Linder for outstanding research assistance. Lucia Alessi and Luca Onorante also thank Sergio Iommi for excellent research assistance. Views expressed are those of the authors and not necessarily those of the European Central Bank, the Federal Reserve Bank of New York or the Federal Reserve System.

[†]Economist, European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany; e-mail: lucia.alessi@ecb.europa.eu

[‡]Bernstein Distinguished Professor of Economics and Professor of Finance, Department of Finance, Kenan–Flagler Business School, and Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; email: eghysels@unc.edu

[§]Head of the Joint Modeling Project, Central Bank of Ireland, Central Bank of Ireland, PO Box 559, Dame Street, Dublin, Ireland; e-mail: Luca.Onorante@centralbank.ie

[¶]Senior Vice President, Macroeconomic and Monetary Studies Function, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045, USA; e-mail: richard.peach@ny.frb.org

^{||}Executive Vice President of the Federal Reserve Bank of New York and Head of the Markets Group, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045, USA; e-mail: simon.potter@ny.frb.org

1 Introduction

The purpose of this paper is to document macroeconomic forecasting by two of the key central banks: the Federal Reserve Bank of New York and the European Central Bank (hereafter respectively FRBNY and ECB). The paper is the result of a collaborative effort between the two institutions, allowing us to study the time-stamped forecasts as they were made throughout the global financial crisis (henceforth referred to as *the crisis*). These time-stamped forecasts of key macroeconomic variables were the inputs of the extraordinary measures taken by central banks both in terms of conventional interest rate policy as well as in terms of the creation of many crisis-driven special purpose facilities which provided much needed support to a financial sector on the brink of collapse.

A number of assessments have been made about the forecast performance of central banks during the Great Recession, including ? and ?. It is well known and well articulated notably by ?, that the forecasting tools and monetary policy processes have been put through a severe stress test by the events of the past five years. Both ? and ? observe that the forecast performance during the Great Recession has been noticeably worse than prior to the crisis.

The analysis in this paper builds on this literature and benefits from a unique access to the time-stamped, real-time forecast records of the Federal Reserve Bank of New York and the European Central Bank. In this respect our analysis is similar and complimentary to ? who provides a thorough review of the Bank of England forecasting performance during the crisis. However, our analysis also goes beyond the traditional forecast evaluation exercises and discusses methodological advances which most likely will shape the future of the macroeconomic forecasting process at central banks.

The paper is organized as follows. In section ?? we start with a brief summary of the institutional backgrounds of the forecasting process at the FRBNY and ECB and we discuss the data used in the study. Section ?? examines the point forecast performance. Since the main trigger of the crisis was a meltdown of the US subprime mortgage market, we pay special attention in section ?? to whether financial market signals were fully accounted for in the central banks forecasting process. Section ?? presents a new conceptual and methodological framework for prediction and policy guidance which grew out of the challenges faced in particular at the FRBNY. The framework has two key ingredients: (1) the emphasis on what might be called scenario-driven forecasting schemes, and (2) the recognition that one should pay attention to distributional features beyond point forecasts, in line with general notions of macroeconomic risk. This analysis naturally links into section ?? which covers model and density forecast evaluations.

2 The Forecasting Process: Institutional Backgrounds and Data Description

Before discussing the specific cases, it is worth recalling some of the literature on the general topic of Central Bank forecasting *prior* to the Great Recession.¹ ? compare the forecast accuracy of the New Area-Wide Model (NAWM), designed and used by the European Central Bank for macroeconomic forecasts, with Bayesian DSGE-VARs, BVARs and reduced-form models. They find that the DSGE-VAR model outperforms DSGE, NAWM, VAR, and BVAR models in forecasting output. Similarly, ? compare the performance of Estimated Dynamic Optimization-based model (EDO) from the Board of Governors of the Federal Reserve Board (FRB) with VAR, BVAR, univariate autoregressive model, and the Greenbook and FRB forecasts. They find that the out-of-sample real time forecasts of the EDO are comparable to the autoregressive model but generally not statistically different from it and to other models.

A second important stream of literature focuses on the forecast of macroeconomic variables using factor models, which are a parsimonious way of extracting large information on overall economic conditions. ?, ? and ? survey applications of these models in forecasting macroeconomic and financial variables. The majority of these studies finds that factor models consistently beat univariate and multivariate autoregressive models and, often, judgmental forecasts both during expansions and recessions. Moreover, ? compares the out-of-sample forecast performance for US output growth of DSGE models with VARs and factor models and finds that factor models generally outperforms any other model in the short run.

A growing bulk of literature focuses on nowcasting and short term forecasting GDP growth using mixed frequency models.² These models are based on the idea that it is possible to use information contained in more timely indicators (available at a higher frequency) to improve the forecast of quarterly output growth. This literature is gaining prominence in particular against the background of the failure of standard models during the global financial crisis, and it will be one of the topics highlighted later in the paper. Early contributions include the use of so-called bridge equations, see e.g. ? and ?, a traditional modeling strategy for obtaining an early estimate of quarterly GDP growth by exploiting information in monthly variables. Using mixed frequency data, ? and ?, among others, have formalized the process of updating forecasts as new releases of data become available((, using the terminology of nowcasting for such updating DELETE?). More specifically, they use state-space models and the Kalman filter as a representation of the joint dynamics of real GDP and the monthly data releases. This allows to nowcast in presence of staggered data releases, a common feature of high frequency indicators, often showing missing observations at the end of the sample due to non-synchronized publication lags (the so called jagged/ragged edge problem). An alternative reduced form strategy based put forward in recent work by ?, ? and ? uses so-called MIDAS regressions. There are some conceptual

¹Excellent recent surveys on the topic appear in ?, ? and ?.

²See in particular the recent surveys of ? and ?.

differences between the two approaches. The former involves latent variable state space models, while the latter is purely regression-based.

Applications of mixed frequency models to nowcast GDP growth in the U.S. or Euro area include ?, ?, ?, ?, ?, ? among others. The general finding is that these models generally outperform nowcasts of GDP growth compared to models that use quarterly frequency only, and are comparable to judgmental forecasts in the US and in the Euro area. In particular, ? find that the nowcasting and forecasting ability of the MIDAS and mixed-frequency VAR (MFVAR) to quarterly GDP growth in the euro area are complements as the MIDAS does better for short horizons (up to 5 months), whereas MF-VAR are better for longer horizons (up to nine months).

3 The Forecasting Process: a comparison

This section highlights the salient features of the forecast conducted at the Federal Reserve Bank of New York and at the European Central Bank. It is worth noting that there are important differences with the FRBNY and the Federal Reserve in general, but such differences are more in the processes than in the ultimate goals and contents of the forecast. First, in the US the forecasting process is event-driven. Namely, the FOMC meeting calendar determines the frequency and occurrence of the forecasting process. In contrast, at the ECB the forecasting process is based on a fixed calendar. There is, however, a continuous nowcasting process which will be further discussed later and also revisited in section ???. Second, the US Federal Reserve focuses on so called core inflation - excluding energy prices, whereas the ECB targets the HICP, which includes such prices. Finally, the ECB is assisted by national central banks that are members of the Eurosystem. As a result, the ECB forecasting process requires a coordination effort between the ECB and the national central banks. In contrast, the FRBNY is a member of the Federal Reserve System but exclusively relies on its own research staff.

At the same time, important common features of the forecasts include the wide use of conditioning assumptions, the involvement of a variety of models, a frequent assessment for reasonableness and consistency of the outputs and a prominent role assigned to expert judgement. Furthermore, both include as a fundamental component the assessment of risks surrounding the projections. On this last point, ((Kilian and Manganelli: Lutz Kilian & Simone Manganelli, 2008. "The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences under Greenspan," Journal of Money, Credit and Banking, Blackwell Publishing, vol. 40(6), pages 1103-1129, 09.)) point out that the same theoretical tools can be adapted easily to the risk management strategies pursued by both central banks.

3.1 Macroeconomic Forecasting at the Federal Reserve Bank of New York

The FRBNY has a long tradition of forecasting US economic conditions. These forecasts and accompanying analysis make up a part of the vast array of information on domestic and global economic and financial conditions which inform the FRBNY's role in the process of setting US monetary policy.

3.1.1 What is the Forecast?

Before beginning to explain how the forecast is prepared, it would be useful to explain what kind of forecast the FED produces. One component of the forecast is a "best guess", based on a set of conditioning assumptions, of the path of macro variables such as GDP growth, the unemployment rate, and inflation over a forecast horizon of about two years. Internally, this best guess is referred to as the "modal" forecast in that it is intended to be the most likely of a wide range of potential outcomes. The modal forecast is not the output of a single model of the US economy. The science of economics has not yet developed a "consensus" model which produces reliable forecasts for a range of key variables. A variety of models of varying degrees of sophistication are used in the process, but ultimately it is a "judgmental" forecast.

The number of underlying conditioning assumptions is extensive and defining them is an important part of the process. The major conditioning assumptions include the economy's potential growth rate, the stance of fiscal and monetary policy, the rate of growth of the economies of the major US trading partners, the path of the exchange value of the dollar, and the path of global oil prices. Putting together this modal forecast is both challenging and humbling. The FRBNY has at its disposal vast amounts of macro and micro data on the US economy covering much of the post-World War II period. In addition, the FRBNY, like many other central banks, has a team of highly trained and seasoned professionals who are able to combine state-of-the-art statistical techniques with the experience and judgment that comes from having monitored and forecasted the US economy for decades. Nonetheless, it is taken for granted that before even the forecasting process starts that the probability that events will unfold as described in the modal forecast is essentially zero. The US economy is constantly evolving. Each cycle is unique and turning points are extremely difficult to identify in real time.

For all of the reasons above, the FRBNY also produces an assessment of the risks around the modal forecast. This risk assessment is based on the construction of the most prominent risk scenarios confronting the global economy, an assessment of the most likely errors in the conditioning assumptions, and the most likely errors in the understanding of current conditions and how the economy actually works. In section ?? we will elaborate on some of the technical aspects.

3.1.2 A High Level Overview of the Process

((I DO NOT WANT TO REWRITE THIS SECTION, AS I AM NOT IN THE TECHNICAL DETAILS, BUT DO WE NEED THE FIRST PARAGRAPH AT ALL? IF SO, WHAT ABOUT PULLING IT OUT AND DESCRIBING US AND EU?))

Most major economies maintain a system of national accounts through which the level and growth rate of economic activity and prices is estimated. In the US, the National Income and Product Accounts or NIPAs are maintained by the Bureau of Economic Analysis (BEA) of the US Department of Commerce. The BEA is for the most part a data user, relying on data produced primarily by the

Bureau of Labor Statistics and the Bureau of the Census. Gross Domestic Product (GDP) is the sum of all final expenditures on goods, structures, and services within the borders of the United States over a given period of time such as a calendar quarter or year. The main categories of final expenditures are consumption, gross private domestic investment, government consumption expenditures and gross investment, and net exports. Data are available to allow each of these major categories to be disaggregated to a quite detailed level. While GDP is the sum of all final expenditures made within the borders of the US over a given period of time, it is simultaneously equal to all of the income which accrues to the factors of production over that same time period. Gross National Product (GNP), or the income earned by factors of production owned by US citizens and corporations, is derived by adding to GDP income flowing into the US from the rest of the world and subtracting income payments made to the rest of the world. Net national product is derived by subtracting from GDP private and government capital consumed (depreciation) during the time period that income accrued. To derive national income, the statistical discrepancy is subtracted from net national product. The statistical discrepancy is, as its name suggests, an accounting identity that equates the income side to the NIPAs with the expenditure side. If the statistical discrepancy is positive, it indicates that the estimated sum of total expenditures exceeds the estimated sum of incomes. If it is negative, estimated incomes exceed estimated expenditures. Finally, national income is composed of, in order of importance, wages and salaries, corporate profits, proprietors' income, etc.

The forecast is assembled in a spreadsheet model of the expenditure and income sides of the NIPA. However, the FRBNY organizes the expenditure side of the NIPAs in a manner consistent with a "bottom up" projection of real final expenditures, the corresponding price indices, and nominal final expenditures. In this alternative presentation, GDP is composed of two main categories of expenditures, domestic demand and net exports. Domestic demand is made up of inventory investment and final sales to domestic purchasers. Final sales to domestic purchasers is divided into two categories, private, which is the sum of consumption and gross private domestic fixed investment, and government consumption expenditures and gross investment. An important metric is the ratio of the stock of inventories over total final sales, which is the sum of final sales to domestic purchasers and net exports.

The forecast process begins by making an initial projection of the quarterly rates of growth of the components of real final sales to domestic purchasers—C plus I plus G—and their corresponding price indices over the forecast horizon. This first pass is based on a judgmental assessment of recent trends, typical cyclical behavior, and exogenous factors such as upcoming changes in tax rates or government spending. To this is added a first estimate of the rate of growth of inventories, keeping the ratio of inventories over final sales on a path judged to be consistent with longer-term trends. The projected growth rate of domestic demand, along with assumptions about global growth and the path of the exchange value of dollar, are then run through a trade model, providing a first estimate of net exports.

The projected path of real GDP growth resulting from this process is then an input into what FRBNY staff calls the "Productivity and Costs" model. This model produces projected paths of the unemployment rate over the forecast horizon. In addition to the growth of real GDP, there are numer-

ous assumptions that must be made for variables such as productivity growth, average weekly hours worked, population growth rate, and labor force participation rate. Finally, the FRBNY runs a "Prices and Wages" model with inputs from the preceding two steps to derive our projections for core CPI and PCE deflator inflation and compensation per hour in the nonfarm business sector.

This set of "first round" results are then assessed for reasonableness and consistency with views about the major forces impacting the economy. Is the path of the personal saving rate consistent with what we know about the state of household leverage and lending standards? Is the path of the unemployment rate and inflation consistent with our assumptions regarding monetary policy and the path of long term interest rates? Is the path of corporate profits consistent with our assumed path of equity prices? Is the path of the inventory/final sales ratio reasonable? Answers to questions such as these often prompt subsequent iterations of the process until a forecast is obtained that can be explained and defended to senior management. Again, as noted earlier, this exercise produces a modal forecast, or the most likely outcome. In section ?? we will examine the statistical accuracy of these point forecasts. In section ?? we elaborate on how various risk scenarios are built around the modal forecast. The latter will produce density forecasts which will be appraised in section ??.

3.2 Macroeconomic Forecasting at the European Central Bank

The forecasting process of the Eurosystem has specific characteristics due to its objectives and to the organization of the European system of Central Banks (ESCB). We first describe the forecasting process in place at the European System of Central banks and then turn to the organization of the forecasting process.

The Eurosystem and ECB staff macroeconomic projections for the euro area are carried out four times a year, twice in the context of the Eurosystem staff Broad Macroeconomic Projection Exercise (BMPE) and twice in the context of the ECB staff Macroeconomic Projection Exercise (MPE). The forecasts are characterized by a special focus on inflation. The quarterly exercises have two main components. The main process, the (Broad) Macroeconomic Projection Exercise, looks at the broad macro-economy, also including fiscal and labor market indicators. The quarterly projections extend to an horizon of two to three years. In parallel and contemporaneously, the Narrow Inflation Projection Exercise (NIPE) looks at monthly movements in inflation up to 18 months ahead. It focuses on all main components of the Harmonized Index of Consumer Prices (HICP): unprocessed and processed food, energy, services and non-energy industrial goods.

The need to coordinate the work of central banks requires the adoption of common technical assumptions underlying the forecasts. The ESCB forecast uses futures for oil prices and a random walk assumption for the exchange rate of the euro. Common food prices assumptions and international environment projections are produced by the ECB staff and adopted uniformly, while National Central Banks (NCBs) update the fiscal policy assumptions on the basis of new passed (or likely to be passed) fiscal measures. The conditioning assumptions for monetary policy are also an important element of

the forecast. Before 2006 an assumption of constant short-term interest rate was used for simplicity and to avoid commitment. Since June 2006, market interest rates are used to ensure a higher consistency of projections and a better comparability with other forecasts while still maintaining a “technical” nature. For this reason, future policy rates are not published in the forecast.

The responsibility of the forecast exercises rests in the hands of the staff of the ECB and that of the NCBs. The process starts by updating the assumptions and providing a first baseline forecast. An iterative process leads to convergence on a new coherent forecast including all the new assumptions; this process includes the NCBs in the June and December exercises and is limited to the different divisions of the ECB in the March and September (MPE exercises). Each forecast is produced contemporaneously via aggregation of the individual Eurozone countries forecasts and as a euro area direct forecast. These two strands of forecasts are continuously discussed and compared and the whole process lasts between 6 and 8 weeks. The NIPE process follows a parallel and largely contemporaneous procedure. The forecast is discussed at the highest level at the ECB and NCBs in subsequent meetings of the Forecast Steering Committee, and the BMPE/NIPE numbers and reports are discussed and endorsed in a final meeting of the Monetary Policy Committee (MPC). Finally, the forecast is presented to the Governing Council on the pre-set publication date.

Ultimately, following general practices and similarly to the FRBNY, the ECB/ESCB forecast is a mixture of models and expert judgement. Notably for (shorter-term) inflation projections a detailed knowledge of institutional factors (e.g. administered prices, indirect taxes, fiscal policy measures) is an essential ingredient. Models used differ according to purpose and they include large-scale macroeconomic models, area-wide models (ECB and some NCBs), the new Multi Country Model (ECB) and NCB models. A wide array of ancillary or more specialized models pertain to the short term (bridge equations, factor models) or deal with specific aspects of the forecast (specialized models for international food commodities or housing market, models of the global economy). An area-wide DSGE model is also used to interpret the forecasts.

The projections are published in the March, June, September and December issues of the ECB Monthly Bulletin and are available since December 2000. The published figures include projections for inflation (as measured by the HICP), the growth of real GDP and its main expenditure components for the current year and the following year (except for the December exercise when the horizon is two years ahead). The projections are published in the form of ranges in order to reflect the degree of uncertainty attached to such exercises. Charts appearing in Figures ?? and ?? report the published (B)MPE projections for respectively Euro area real GDP and HICP in terms of ranges for the 2000-2013 sample. The solid horizontal lines across each year denote the actual outcomes. For each target year, a total of nine forecast releases are published, with the first published in the December exercise two years prior to the target year and the last published in the December Monthly Bulletin of the target year. The midpoint of the range is also indicated for each of the nine prediction exercises.

One point that should be clarified is that the ranges reported in Figures ?? and ?? are not derived by mechanically considering the point forecasts yield by a suite of models as one might think. The

models used in the (B)MPE are very many as not only the ECB uses a suite of different tools, but also many of the national central banks do. Some models are designed to forecast other variables than GDP or inflation, e.g. oil prices or housing prices, and may yield GDP and inflation forecasts only as a by-product. Country models are also in some cases different from each other, reflecting national specificities. Moreover, hard and soft data also inform the final forecast (range), while not necessarily entering any models.

Risk analysis plays in general an important role in central banks; at the ECB, the risk analysis reflects the view of the MPC, whose assessment is complemented by model-based estimations of uncertainty. Risk factors typically discussed include whether oil prices will follow futures prices, the future dynamics of the exchange rate, risks to the global growth projections, and risks stemming from wages, housing markets or unforeseen fiscal policy measures. Given the importance of price stability in the ECB's mandate, risks to price stability are additionally discussed.

The need to take into account the monthly meeting frequency of the ECB's Executive Board leads to an additional process of constant updates of the projections between exercises. Price updates are produced by the ECB staff after each release of the flash estimate for HICP at the beginning of the month and in the middle of the month when final figures for the previous month become available. New figures are produced by updating the most recent projections to take into account the publication of additional data and the changes in the assumptions (with a particular emphasis on international energy, on updates of food prices and exchange rates) and in other information available at higher frequencies (e.g. weekly oil bulletin). These projections are of a mechanical nature, to account for the fact that the judgement is already included in the quarterly exercise.

((LUCIA CHECK THIS IS CORRECT)) On the real side of the economy a similar process of update produces short-term mechanical early estimates for quarter-on-quarter euro area real GDP growth twice a month, namely after the release of industrial production data in the middle of the month and following the release of EC survey data at the end of the month. For each quarter, therefore, a sequence of twelve estimates is produced for GDP growth in a certain quarter t , two per month in six consecutive months, starting in the second month of the previous quarter $t - 1$ and ending in the first month of the next quarter $t + 1$ (see the upper panel of Table ??). For instance, the first forecast for GDP growth in the second quarter of the year is made in February, while the last estimate is done in July. The GDP flash estimate for the second quarter is released in mid-August. The schedule appears in Table ??. This process is complemented by the use of bridge and factor models for exploiting real-time information on GDP ((ARE HE LATTER EARLY ESTIMATES?)).

Real-time GDP forecasts (so called GDP Early Estimates - henceforth EE) are part of this constant updating as well. Contrary to the BMPE case, these real-time forecasts are produced on a mechanical basis and therefore it is possible to provide a more detailed description of the models used in the framework of the Early Estimates. The short-term forecasting exercise is described in the article *Short-term-forecast of economic activity*, ECB Monthly Bulletin, April 2008, pp.70-74. The models used to produce them are based on systems of bridge equations (SBE) or dynamic factor models (DFM)

and on one of the following approaches: (i) a direct approach (DIR), providing forecast for euro area GDP directly, as opposed to an aggregation of forecasts; and (ii) an approach aggregating value added (VAD). The models are described in the following ECB Working Papers (between parenthesis published versions): no. 1189 (?), no. 1051, 2009 (?), ECB Occasional Paper no. 84, ECB Working Paper no. 953 (?). For the class of dynamic factor models see also ECB Working Paper no. 949 (?) [DFM-DIR], no. 975, [SBE-VAD], no. 622 (SBE-DIR), and no. 276 (SBE-DIR).

4 Evaluation of Point Forecasts

In this section we start with a relatively conventional forecast evaluation: examining point forecast performance. It should be pointed out that the purpose of this exercise is not to run a horse race, as comparing the performance between the ECB and FRBNY is extremely difficult. First of all, as noted in section ??, the institutional details of the forecasting process at the ECB and FRBNY are very different. This poses some challenges for a direct comparison; in particular, the fact that the ECB forecasting process is almost entirely calendar-driven, whereas the Federal Reserve forecasting process is FOMC meeting-driven make the outcomes difficult to compare as they are conditioned on different information sets. For similar reasons, it is also not so straightforward to compare central bank forecasts with those of professional forecasters. In fact the results are mixed in this regards. ? observes that the Bank of England forecasts have been marginally worse than those of outside forecasters. In contrast, Figure ?? reported by ?, which compares the April 2008 assessment of the downside risk in the New York Fed research staff outlook with that of the Survey of Professional Forecasters, shows there is some evidence that FRBNY had a more sober assessment of the risk of a severe downturn than did private sector forecasters. We will elaborate more on this in section ??.

Since the crisis started in the US, it is natural to start with the FRBNY forecasts. ? identifies three main failures in the real-time forecasting of the FRBNY which stand out:

- The misunderstanding of the housing boom: the FRBNY staff analysis of the increase in house prices did not find convincing evidence of overvaluation. Perhaps the best illustration of this is provided in Figure ?? showing the vintage FRBNY housing starts forecasts during the Great Recession.
- A lack of analysis of the rapid growth of new forms of mortgage finance.
- The insufficient weight given to the powerful adverse feedback loops between the financial system and the real economy.

We focus first on real GDP growth forecasts - the series common to the ECB and FRBNY since the measures of inflation considered differs across the Atlantic. Recall that the ECB projections are published in March, June, September and December. We focus on the December forecast prior to the target year. This is referred to as release five, and hence the middle of the nine vertical bars depicted in

Figure ?? for each of the years in the sample. This particular release is picked as it is the best match with the end-of-year FOMC forecasts prior to the target year.

Table ?? shows one-year ahead ECB and FRBNY forecasts of real GDP growth. Errors are computed as predicted minus realized. The table further reports several accuracy measures: mean forecast error (MFE), mean squared forecast error (MSFE), mean absolute forecast error (MAFE), root mean squared forecast error (RMSFE), minimum forecast error (MIN) and maximum forecast error (MAX). The entire period covers 2000-2012, the *Pre-Crisis* sample covers 2000-2007Q4 and the remainder of the sample corresponds to the crisis period.

The results in Table ?? indicate that the MFE for the ECB and FRBNY are quite comparable, with the ECB featuring slightly lower numerical values. The errors for GDP growth are on average positive, meaning that growth was overestimated, with the wedge being larger during the crisis. In terms of mean squared forecast error, mean absolute forecast error, root mean squared forecast error, minimum forecast error and maximum forecast error the results are even more similar. Note that during the crisis the MSFE for the ECB and FRBNY increased respectively six and five times in comparison with the pre-crisis sample, and the MAFE and RMSFE more than doubled. This may seem a relatively poor record for both central banks, but the lower accuracy is not particularly exceptional when compared for instance with the Bank of England record reported in ?.

The picture is quite different for inflation, however, as is evident from Table ?. In terms of MSFE and MAFE, while we see comparable numbers in the pre-crisis sample for the ECB and FRBNY, there is a clear divergence during the crisis. The FRBNY record on inflation forecasting record during the crisis was identical to its pre-crisis performance. In all cases the errors are on average negative, meaning that predicted inflation was lower than realized. For the US the mean during the crisis period is for all practical purposes zero. The ECB's record in contrast worsened significantly. The MSFE and MAFE of inflation forecasts became two to three times larger for the ECB. It is important to recall that both institutions focus on different inflation measures, since the ECB targets an index which includes energy prices. In particular, the high volatility of food and energy prices, included in the ECB projections, is at least partially responsible of the decrease in the ECB's forecasting performance . It would therefore be more appropriate to compare the performance of the ECB inflation forecasts with that of the Bank of England. Indeed, ? shows that the British central bank inflation forecasts also showed substantial deterioration, roughly of the same order of magnitude. In Table ? - discussed later - we report the forecast performance for the GDP deflator and CORE GDP deflator of the FRBNY. We observe that the FRBNY forecast performance of the GDP deflator is also considerably worse during the crisis.

In section ?? we discussed the fact that the FRBNY and the ECB have some significant operational differences in their forecasting process. As a result, the historical records that are available differ as well. In the remainder of this section we take advantage of the specific features of each central bank to highlight topics of interest.

Recall that the ECB follows a fixed calendar forecasting process. The advantage of such a setup is that we can easily appraise fixed-horizon forecasts. In Tables ?? and ?? we limited our assessment to

the one-year ahead horizon only, corresponding to release 5 for the ECB, since it was the easiest for the purpose of comparing both central banks (the Federal Reserve holds a December FOMC meeting). The ECB produces 9 forecasts for a given target year, and Table ?? now reports the summary statistics for three different releases: 1, 5 and 9, where the former corresponds to a two-year horizon and the last is the end-of-year forecast of the target year. The reported statistics are the same as those in the previous tables. We observe that the deterioration is across the board. For example the MSFE of release 1 pre-crisis was 2.44 and increased to 17.93 during the crisis. Interestingly, releases 1 and 5 look very similar, prior and during the crisis. Release 9, which is the last forecast in December of the target year had a minuscule MSFE of 1.10 prior to the crisis which increased five-fold to 5.74. The RMSFE for that release is 2.40, meaning that the ECB overestimated GDP growth by 2.40 % during the crisis - or more accurately failed to estimate the depth of the recession by the quoted percentage.

Table ?? is specific to the FRBNY and sheds light on the forecast performance for the various components of GDP. Recall that the modal forecast is assembled in a spreadsheet model of the expenditure and income sides of the NIPA and an alternative presentation, where GDP is composed of two main categories of expenditures, domestic demand and net exports. We concentrate again on the one-year horizon. The entries to the table are Real GDP (now in terms of relative forecast errors), PCE, BFI ETS (Business Fixed Investment, Producers Durable Equipment), BFI Struc (Business Fixed Investment, Nonresidential structures), Residential Investment, Exports, Imports, Federal Government, Gov S+L (state and local), PCE deflator and Core PCE deflator. From the table we note that during the crisis, Business Fixed Investment, Producers Durable Equipment, Business Fixed Investment, Nonresidential structures, Residential Investment, Exports, Imports and PCE deflator all featured substantial forecast errors. These results are not surprising. The collapse of the housing market, and the subsequent impact on the real economy were unprecedented and off the charts.

5 Were Financial Market Signals Fully Accounted For?

Economic forecasters, like most other economists, were behind the curve as the global financial crisis developed. Both theoretical models, as well as empirical ones, either ignored or failed to take fully account of the powerful adverse feedback loops between the financial system and the real economy.

The question we address in this section is whether central banks could have done a better job at reading financial market signals. The crisis was first and foremost a crisis caused by financial market turmoil. The analysis in this section is influenced by ? who suggest to rely more on financial market data to improve macro forecasting. The issue requires one to think about mixed frequency data since financial market data are intrinsically high frequency (in our case daily) whereas the forecasts pertain to low frequency macroeconomic phenomena such as GDP growth. One solution, is to use MIDAS regression models, adopting a framework put forward in recent work by ? and ?.³

³Recent surveys on the topic of MIDAS include: ? who review more extensively some of the material summarized in this document, ? who provide a very simple introduction to MIDAS regressions and finally ? who discuss volatility models and

We start with the real-time GDP Early Estimates - or EE - of the ECB. Recall that these short term mechanical early estimates for quarter-on-quarter euro area real GDP growth are produced twice a month, namely after the release of industrial production data in the middle of the month, and following the release of EC survey data at the end of the month. For each quarter, therefore, a sequence of twelve estimates is produced for GDP growth in a certain quarter t , two per month in six consecutive months, starting in the second month of the previous quarter $t-1$ and ending in the first month of the next quarter $t+1$. In the lower panel of Table ??, we depicted the variables which enter the MIDAS regressions. Namely, we want to study whether the, respectively forecast errors e_t^i $i = 1, \dots, 4$, nowcasts e_t^i $i = 5, \dots, 10$ and backcast errors e_t^i $i = 11$ and 12 (cfr. top panel of Table ??) are predictable with real-time financial data available at the time the fore/now/backcast is made - meaning the end of every other week. We use the convention that $e_t^i \equiv EE_t^i - \Delta GDP_t$. We denote the daily financial series available at the time of the EE release by HF_t^i , as depicted in the lower panel of Table ?. There are in fact $j = 1, \dots, K$ high frequency series. Moreover, in the middle of quarter t a first release of ΔGDP_{t-1} is made. This means that e_{t-1}^i is observable for $i = 8, \dots, 12$, and only e_{t-2}^i is available for the first seven, yielding the following autoregressive scheme can be used in the MIDAS regressions, using the vintage i of the j^{th} daily financial series denoted $HF_{j,t}^i$:

$$\begin{aligned}
e_t^i &= \mu_j^i + \rho_j^i e_{t-2}^i + \beta_j^i \sum_{d=0}^{D_i} w(\theta_j^i) HF_{j,t_i-d}^i + u_{jt}^i, & i = 1, \dots, 7 \quad j = 1, \dots, K \\
e_t^i &= \mu_j^i + \rho_j^i e_{t-1}^i + \beta_j^i \sum_{d=0}^{D_i} w(\theta_j^i) HF_{j,t_i-d}^i + u_{jt}^i, & i = 8, \dots, 12 \quad j = 1, \dots, K \quad (5.1)
\end{aligned}$$

The time subscripts to the HF refer to daily lags $t_i - d$ where t_i is the day of EE release i and d are daily lags. While there are several possible parameterizations of the MIDAS polynomial weights, to keep the parameter specification tight we take a restricted Beta polynomial with only one parameter, see ? or ?. In other words, we are testing the assumption that financial data contain information which is useful for explaining early estimates forecast errors, because it is not contained in low(er) frequency macroeconomic variables. The financial variables we consider belong to the following five groups: fixed income, foreign exchange, commodities, stock exchange, corporate yields. The detailed list of the indicators is provided in Appendix ?? and the real-time data schedule appears in the lower part of Table ??.

For each t we have $j = 1, \dots, K$ predictions of the fore/now/backcasts e_t^i . The next step, following the strategy put forward in ? is to combine the K predictions. ? provides an excellent survey of forecast combination methods. We consider two strategies: (1) simple average across the MIDAS regressions each involving one of the financial series considered and (2) Bayesian model averaging (BMA). The latter involves a weighting scheme parameter α which encompasses the BMA when ($\alpha = 1$), the simple average of forecasts ($\alpha = 0$) and Bayesian Model Selection (BMS), in which only the best model is used at each point in time (for $\alpha \rightarrow \infty$). The new parameter can be interpreted as a filter for the

mixed data sampling.

number of models to be included in the forecasting combination: for low values all the models are included, increasing values of α imply a progressive tighter selection of the models, to the extreme of choosing only the best one. Using this additional degree of freedom, we can provide a forecast combination based on an optimal α , simply by specifying a prior distribution for α and letting the data determine this degree of tightness. A non informative prior for α was used in the exercise.

The results are reported in Table ???. We only cover the simple model averaging, as the BMA results yielded similar findings. The sample runs from 2008Q2 to 2012Q4. The table reports the average R^2 across all MIDAS models estimated for each of the releases. The benchmark model is a simple autoregressive model, i.e. equation (??) without the high frequency data. In order to test the significance of the results, we built 95 % confidence intervals for the R^2 of the autoregressive model by bootstrapping its errors.⁴

We do not report all the detailed results, only summary statistics. Namely, for each release we report the average R^2 for the aforementioned MIDAS regressions. In addition, we also report the same statistics for the autoregressive models which exclude the financial series. The former R^2 s are typically around 30 % with some variation across the releases. In contrast the autoregressive models typically have a 5 % R^2 . The last entry in the table reports the bootstrap intervals for the autoregressive model. All the MIDAS regression R^2 exceed the upper 95th percentile, implying that the MIDAS regressions yield significantly higher R^2 s. Omitting the details of the individual series, we can summarize the findings as follows:

- The forecast error variance explained by the single best MIDAS model ranges from 50 % (release 12) to more than 70 % (releases 5, 6 and 7).
- Fixed income variables, especially the shorter end of the curve, turn out to be among the best explanatory variables across all releases, followed by commodities and the VSTOXX (an implied volatility index similar to the VIX).
- The EONIA (Euro OverNight Index Average) performs particularly well in predicting the forecast error of the first three releases (forecast), while for later releases (mainly nowcast) the Euribor, in particular the 6-month, yields the best results.
- The performance of the VSTOXX is particularly good for the fourth release and tends to progressively worsen with subsequent releases.
- Commodities display the largest explanatory power for the errors made when nowcasting GDP growth towards the end of the current quarter (releases 9 and 10) and backcasting.

⁴Adding regressors necessarily raises the R^2 . Hence, in order to compare whether the increase in the R^2 due to the inclusion of high-frequency regressors on top of the own lag is not spurious one should refer to the adjusted R^2 , which penalizes the inclusion of additional explanatory variables. The qualitative results presented, in particular the magnitude of the improvement yield by the high-frequency regressor, do not change when considering the adjusted R^2 . For ease of interpretation, we stick therefore to the R^2 . The number of bootstrap repetitions is 1000.

- With respect to corporate spreads, both AAA-BBB and AAA-government bond spreads are generally among the variables yielding non-significant or only marginal improvements over the benchmark.
- MIDAS regressions including the exchange rate always yield a poor R^2 .

We also computed MIDAS regressions for the Eurosystem staff Broad Macroeconomic Projection Exercise (BMPE), focusing on the fifth release. The MIDAS regressions average R^2 for GDP forecast errors for the fifth BMPE covering the sample 2000Q1 - 2012Q1 is 0.45, whereas the bootstrap confidence interval for the autoregressive model is $[0.00, 0.26]$, yielding the same type of conclusions. The other BMPE releases are similar. Note that the BMPE sample does not only cover the global financial crisis, although the sample is much smaller compared to the EE.

The overall conclusion is that financial variables contain useful information for predicting euro area real GDP growth in real time. Simple MIDAS regressions using such financial series could have improved the forecast accuracy.

We also examined whether some classes of models performed better than others. Since the EE forecasts are mechanically generated by a select class of models without judgemental corrections we can compare the performance of the separate classes. We noted in section ?? that the models used to produce the EE are based on systems of bridge equations (SBE) or dynamic factor models (DFM) and on one of the following approaches: (i) a direct approach (DIR), providing forecast for euro area GDP directly, as opposed to an aggregation of forecasts; and (ii) an approach aggregating value added (VAD). We do not report any of the results explicitly since the finding was rather straightforward to summarize: all classes of models did roughly the same over the sample period considered.

In Table ?? we run the same type of analysis with *all* the FOMC forecasts of the FRBNY using financial data described in Appendix ?. The dates are each FOMC meeting between early 2002 and the end of 2012. Recall also that the *Pre-Crisis* sample covers 2000-2007Q4 and the remainder of the sample corresponds to the crisis period. The series we consider roughly match the realm of financial instruments used in the euro area exercise, although the number of series is smaller. Note also that the setting is somewhat different, since the forecasts we consider here are the modal forecasts, not mechanically generated by models but rather subjected to a process involving judgemental corrections. The conclusions are the same as for the euro area. Note that we are also able to do a sample split before and during the crisis. While sample sizes are small, there is evidence that the usefulness of financial series increased during the crisis, when one compares the regression fits.

6 The Forecasting Process: A Conceptual Framework

As noted earlier, research departments at central banks typically run a host of models. These typically range from simple bridge equation models, univariate time series models, MIDAS regressions, to

multivariate models such as VARs, factor and DSGE models. These produce a range of forecasts as was illustrated in Figures ?? and ??. The recent global financial crisis pushed central bank forecasters into uncharted territory. Models exclusively relying on historical estimates posed many challenges to forecasters. In this section we present a framework that reflects some of the new thinking prompted by the challenging new environment faced by central banks. The framework has two key ingredients: (1) the emphasis on what might be called scenario-driven forecasting schemes, and (2) the recognition that one should pay attention to distributional features beyond point forecasts, in line with general notions of macroeconomic risk.

The first point is not exclusively related to macroeconomic forecasting exercises. Central banks started a process of regular stress tests for the financial sector in the aftermath of the crisis. The results of a first comprehensive, forward-looking assessment of the financial conditions of the US 19 largest bank holding companies by the federal bank supervisory agencies were released early May 2009. The exercise was conducted by the Federal Reserve, the Office of the Controller of the Currency, and the Federal Deposit Insurance Corporation. The Supervisory Capital Assessment Program, publicly described as the bank stress tests, used two macroeconomic scenarios, one based on baseline conditions and the other with more pessimistic expectations, to plot a 'What If?' exploration into the banking situation in the rest of 2009 and into 2010. Baseline scenario assumptions for real GDP growth and the unemployment rate for 2009 and 2010 were assumed to be equal to the average of the projections of the recession's likely depth and duration published by Consensus Forecasts, the Blue Chip survey, and the Survey of Professional Forecasters in February 2009. This baseline, labeled the "adverse scenario" was intended to represent a consensus view about the depth and duration of the recession. The more adverse scenario was constructed from the historical track record of private forecasters as well as their current assessments of uncertainty and amounted for example to a 3.3 % decline in GDP in 2009 measured on a year over year basis (versus -2 baseline decline) and modest 2010 recovery with .5 % growth (versus 2 % baseline). The subjective probability assessments provided by participants in the 2009 January Consensus Forecasts survey and the February Survey of Professional Forecasters implied a roughly 15 percent chance that real GDP growth could be as least as low, and unemployment at least as high, as assumed in the more adverse scenario. This type of analysis, exemplified by the bank stress tests also figured prominently in macroeconomic forecasting and will be further discussed in this section.

The second point is not exclusively related to the crisis. Linear models with certainty equivalence properties where point forecast rules are sufficient (i.e., expected values) are a commonly used framework to discuss policy decisions. The convenience and simplicity of such a framework is widely recognized, but it is also gradually being replaced by settings where macroeconomic risk and uncertainty play a more explicit and prominent role. Several authors, including ?, ?, ?, and many others, have stressed that, when decision makers face uncertainty about the state of or the model for the economy and have a loss function that factors in an aversion to this uncertainty, it can be optimal for a central bank to react to changes in the perception of macroeconomic uncertainty.

It is the purpose of this section to describe some of the attempts that are made at the New York Fed

to incorporate risk and uncertainty into the decision making process and in particular as it pertains to the recent global financial crisis.⁵

6.1 Scenario-driven Risk Profiles

To set the stage, we let \mathbf{Y}_t be a vector containing J observable economic variables, e.g. GDP growth, core inflation, etc. We are interested in thinking about future values of $\mathbf{Y}_{T+s} = \{Y_{jT+s} : s = -\tau, \dots, 0, 1, \dots, S; j = 1, \dots, J\}$, where T is the current time. We let $F_T(Y_{jT+s})$ be a marginal or conditional cdf of the joint cumulative forecast distribution, where: $F_T(-\infty) = 0, F_T(\infty) = 1, F_T(y') - F_T(y) \geq 0$, if $y' \geq y$. At some future finite date s' $Y_{T+s'}$ is known: $F_{T+s'}(y') = 1$, and $F_{T+s'}(y) = 0$ otherwise.

We noted in section ?? that one component of the forecast is a "best guess" (modal forecast), based on a set of conditioning assumptions, of the path of GDP growth, the unemployment rate, and inflation over a forecast horizon of about two years. In addition the FRBNY also makes an assessment of the risks around the modal forecast. This risk assessment is based on judgments of the most prominent risk scenarios confronting the global economy, an assessment of the most likely errors in the conditioning assumptions, and the most likely errors in the understanding of current conditions and how the economy actually works.

We generalize the modal forecasting framework by considering $K + 1$ unobserved states, with state 0 being a baseline state also referred to as a central scenario or modal forecast. More specifically, let

$$\mathbf{Y}_{T+s} = \sum_{k=0}^K 1(Z_{T+s} = k) \mathbf{X}_{T+s}(k),$$

where Z_{T+s} is a stochastic process that takes the values $0, 1, \dots, K$ and $\mathbf{X}_{T+s}(0)$ is random vector produced by the modal/baseline forecast and is Gaussian/symmetric:

$$\mathbf{X}_{T+s}(0) \sim MN(\mu_{T+s}, \Sigma_{T+s})$$

with $\{\mathbf{X}_{T+s}(k), k = 1, \dots, K\}$ having densities f_{T+s} with the property, $\mu_{T+s} = \arg \max_{\mathbf{x}} f_{T+s}(\mathbf{x})$ implying that modes of the individual densities are the baseline forecasts.

If we add transition probabilities across the $K + 1$ states we end up with a Markovian mixture of (truncated) normals. Clearly, some discipline needs to be imposed to avoid parameter proliferation in this model framework and allow relatively direct expression of prior beliefs. The approach is a generalization of the famous fan charts produced by the Bank of England starting in the late 1990s. Unlike the Bank of England approach based on a two-piece normal distribution, the state of the stochastic process Z_{T+s} is associated with a specific economic scenario with the assumption that the baseline or modal

⁵The material in this section is based on material presented by Simon Potter at the Bank of England, October 2010.

scenario was realized at some point in the past and is an absorbing state in the long run. The stochastic process is assumed to be Markovian with a non-homogenous transition probability matrix, with the property that after a move out of the baseline (in this example into a recession or boom), any transition back into the baseline is absorbing:

$$\begin{array}{l}
 \text{Boom} \\
 \text{Baseline: Normal Growth} \\
 \text{Recession}
 \end{array}
 \left[\begin{array}{ccc}
 p & 1-p & 0 \\
 \left\{ \begin{array}{l} 0 \\ pq' \end{array} \right. & q = \left\{ \begin{array}{l} 1 \\ q' \end{array} \right. & nq = \left\{ \begin{array}{l} 0 \\ nq' \end{array} \right. \\
 0 & 1-n & n
 \end{array} \right].$$

Note that the parametrization of the transition matrix is greatly simplified by assuming that the baseline state is absorbing in the long run. The justification for this simplification is that the central bank is assumed to achieve its goals over the long-run by use of its policy instruments and the multivariate gaussian distribution associated with state 0 captures the long run features of such a situation. Further, uncertainty about the current state of the economy can be allowed for by assuming that the stochastic process Z_{T+s} took the value 0 a certain number of quarters before.

The FRBNY implementation of the general $K + 1$ state model maintains this parsimonious transition scheme. Further, rather than make the heroic assumption that the transition probabilities are known, Dirchlet distributions are used to allow for uncertainty and ease updating of beliefs in real time. Thus, the approach essentially amounts to creating random transition matrices with some discipline imposed, most importantly: $\lim_{s' \rightarrow \infty} P[Z_{T+s'} = 0] = 1$. Transitions from the central scenario are drawn from Dirichlet distribution where beliefs about the likelihood about certain scenarios are crucial and degrees of freedom of the distribution is used to capture confidence in these beliefs. For transitions back to the central scenario one needs to chose $2K$ additional parameters. Again this can be parameterized such that one parameter is informed by beliefs about the average time spent in this scenario and another parameter confidence in this belief. Finally, while in principle one could allow transitions between scenarios, they are by design set equal to zero, although it is possible, in general setting, to fill out more of the transition matrix.

Realizations of this process are stochastic sample paths with random entry into the different scenarios and conditional on entry, time in any scenario is random as well. As a consequence there is no unique way of defining the length or time of arrival into a scenario. The definition adopted by the FRBNY implementation of a path associated with a scenario k is the collection of paths which visit at least once the k^{th} discrete state. One can then compute averages across all paths with at least one visit into the scenario. The obvious advantage that it is easy to compute such summary statistics. Yet, the definition also has the disadvantage that it combines paths from central and alternative scenarios. An alternative would be to weight by the occupation times for each scenario.

Simulations are a natural and convenient way to construct forecast distributions from this Markovian mixture. One such simulation works by drawing standard normals for each horizon then:

- For central scenario the scale is determined by the relevant Cholesky factorization of the covariance matrix
- For others scenarios take absolute value of Gaussian draws, then change sign to positive or negative depending the characterization of the scenarios
- Scale either deterministically and sometimes by random variable (i.e., inverse chi-squared for t distribution) to generate fat tails if required

A practical example, inspired by the events during the global financial crisis is as follows. Consider the joint distribution over output and inflation where one can think of a range of scenarios with random draws as follows:

1. Productivity boom: draw + output, - inflation
2. Loss of Credibility: draw - output, +inflation
3. Overheating: draw large +inflation, + then - output
4. Global Credit Crunch: draw large - inflation and output

Before proceeding with the details, particularly regarding the choice of parameters, it is worth visualizing the output that is produced from the above exercise. The two panels of Figure ?? display the sample paths associated with end of 2009Q3 projections of the aforementioned scenarios, namely in addition to the baseline scenario we have: (a) Productivity boom, (b) Loss of Credibility, (c) Overheating and (d) Global Credit Crunch. The simulated paths in Figure ?? feature patterns which indeed correspond to the assumed scenarios. For example *Loss of Credibility* results in rising inflation - more so that the in the scenario *Overheating*, whereas the *Global Credit Crunch* path enters deflationary territory. Likewise, GDP growth is the highest under the *Productivity boom* scenario and the smallest in the *Global Credit Crunch* one, where the latter is projecting negative or zero GDP growth until the end of 2011.

Each scenario path is computed as an average across paths with at least one state visit associated with the scenario. The simulation also produce a forecast distribution as displayed in the panels of Figure ?? where the collection of all drawn paths appears. The central scenario appears, as expected, at the center of the fan chart. The shades of the chart reflect the percentiles of the forecast distribution. These forecast distributions will be the subject of analysis in the next section.

6.2 Calibration

Where do the probabilities of the different states come from? And, how are the parameters of the state-dependent schemes determined? In an abstract sense it is assumed that the policymaker knows/gives

probabilities, modal forecast and history produces the other parameters. In practice, introspection and iteration take place, where introspection can include comparing to historical distributions. In this subsection we elaborate on the calibration of the simulated scenario-driven paths displayed in Figure ??, or any generic implementation.

To visualize which probability schemes are involved in the calibrations, consider the histograms displayed in Figure ?? The left panel provides the probabilities pertaining to the central scenario, namely the probability of remaining in the scenario through 2010Q4.

Broadly speaking four types of calibration methods are used. They can be characterized as: (a) Financial Markets Data-based Calibration, (b) Professional Forecasters Data-based Calibration, (c) Historical Event-based Calibration and finally (d) DSGE Model-based Calibration. In this section we cover the first three (see ? for a general discussion which includes DSGE Model-based Calibration).

The first calibration method can be used to choose a subset of parameters by calibrating to inferred financial market moments. The latter obviously requires a mapping between forecast distribution and policy rates. A good example is the *Loss of Credibility* scenario appearing in Figure ?. Each scenario-driven path implies a Fed Funds rate profile. For the purpose of comparison the plot also displays the market projections extracted from the Fed Fund Futures at different maturities.

A prominent example of Professional Forecasters Data-based Calibration was already mentioned in the context of bank stress tests where the baseline scenario assumptions for real GDP growth and the unemployment rate for 2009 and 2010 were assumed to be equal to the average of the projections published by Consensus Forecasts, the Blue Chip survey, and the Survey of Professional Forecasters in February 2009. There are limitations here as well. Typically professional forecasts provide little path information and often have a year over year format which is not so easy to match with calibrations.

Given the focus on the crisis it is worth elaborating on calibration based on historical events as past financial crises produced actual and hypothesized/feared behavior outside of the standard post-war experience. Simulation allows for construction of probabilistic information on various big events such as: (a) a Great Depression and its associated size of output drop and scale of deflation, (b) stock market crashes and depressions as well as the (c) scale of recoveries after big recessions and (d) the scale of recovery after financial crises. Figure ?? and ?? display conditional CDFs of objects of interest relative to history, namely: depth of recession and historical recovery patterns.

In Figure ?? we consider the probability distribution of the four consecutive quarters with the lowest GDP growth in a recession. The top panel represents a calibration by the FRBNY staff done in April 2008, with a 80 % probability of recession. The calibrated the probability distribution of the four consecutive quarters with the lowest GDP growth in a recession is plotted against historical benchmarks, the 1990-1991 recession with a -1.0 %, 1981-1982 with -2.7 % and a post Great Depression observation from 1937-1938 with a -10 % drop. In the lower panel of Figure ?? we plot a calibration before and after the Lehman failure. The solid line corresponds to a November 20, 2008 calibration (and therefore after Lehman) based on a 97 % probability of recession (even in November 2008 the simulation had

some paths that did not produce a recession), whereas the dashed line is based on beginning of August 2008 calibration with a 46 % of recession. The central forecast of -2 % increases from a less than 20 % probability event to a more than 60 % one. As can be seen the shape of the conditional CDF also changes as more weight is placed on extreme scenarios in the underlying Markov process and the fat tail of the extreme scenarios is increased. The actual depth of the 2007-09 recession was - 5 % (using the pre-2013 benchmark data). As gauged by this metric the August 2008 calibration attributed less than 3 % probability to the ultimate outcome, and it was only by November 2008 that the probability of the actual outcome was close to 15 percent.

Figure ?? displays the conditional CDF of recovery relative to history with the top panel covering a calibration scale of recovery through the end of 2010 with a probability of trough by end of 2009 which was assessed in May 2009 as 73 % and in October roughly the same at 75 %. The probabilities are benchmarked against historical precedents. To lower panel expands the horizon by one year displaying the scale of recovery through the end of 2011 with a probability of trough by end of 2010 which was assessed in May 2009 as 95 %, again measured against historical precedents.

7 Model and Density Forecast Evaluations

In the previous section we explained how to formulate and compute scenario-driven paths and noted that the simulations also produce a forecast distribution as displayed in the panels of Figure ?. In this section we compare these densities with Survey of Professional Forecasters data. In particular, we consider the cross-section of point forecasts of the SPF against the scenario-driven densities produced by the FRBNY throughout the crisis. The purpose of the exercise is to see how extreme views as well as consensus forecasts in the cross-section of the SPF compare with the FRBNY densities. Are the pessimists in the SPF completely out of sink with the FRBNY? Note that we compare point forecasts with density forecasts and this is done on purpose. What we try to address is how far in the tails are the most optimistic and pessimistic professional forecasters? We do this via Q-Q plots of the cross-section of SPF against the FRBNY density forecasts. In Figure ? we plot at yearly intervals such Q-Q plots in October 2007 through 2012. Hence, the plots give us yearly snapshots throughout the global financial crisis how the cross-section of SPF measures up against the plausible scenarios considered by the FRBNY. The left tail of the SPF consists of the most pessimistic point forecasts and the right tail the most optimistic ones. In October 2007 the pessimists lined up with the left tail, and the optimists were not particularly outliers in the right tail. Hence, at the outset of the global financial crisis we note that the cross-section of the SPF and the FRBNY scenarios were well aligned. This completely changes in 2008 and 2009. The cross-section of SPF is overall more optimistic than the FRBNY. Gradually, as the US economy started to slowly recover over the next set of years we see that the SPF become less optimistic when measured against the range of outcomes considered in the simulated paths of the FRBNY models. Figure ? is fascinating. It suggests that the FRBNY appears to be quicker than the collect of point forecasts in the SPF in recognizing the downside risk of the crisis. As the crisis

unfolded, it also appears that the SPF showed more inertia in adjusting its range of forecasts.

INCOMPLETE

Technical Appendices

A Euro area financial data details

The following financial data have been used for estimating MIDAS models for the euro area. The indicators in bold are available since 2000, therefore could be used to explain the Eurosystem/ECB staff forecast errors, while the others are available only for the latest years and have been used only for explaining the early estimates forecast errors.

1. Fixed income

Eonia rate (EONIA); Euribor 1-month (EURIBOR1M); Euribor 3-month (EURIBOR3M); Euribor 6-month (EURIBOR6M); Euribor 1-year (EURIBOR1Y); Euribor Spot week (EURIBORSW)**Euribor Spot week (EURIBORSW);** 3-month Euribor-Eonia swap (EURIBOR_EONIA_SW); European Union 2-year Government Benchmark bond yield (GOVBOND2Y); European Union 3-year Government Benchmark bond yield (GOVBOND3Y); European Union 4-year Government Benchmark bond yield (GOVBOND4Y); European Union 5-year Government Benchmark Bond yield (GOVBOND5Y); **European Union 6-year Government Benchmark bond yield (GOVBOND6Y);** European Union 7-year Government Benchmark bond yield (GOVBOND7Y); European Union 8-year Government Benchmark bond yield (GOVBOND8Y); European Union 9-year Government Benchmark bond yield (GOVBOND9Y); European Union 10-year Government Benchmark bond yield (GOVBOND10Y); **Euribor 1-year Euribor 1-month spread (SPREAD_1Y_1M);** 10-year Government bond yield - Euribor 3-month spread (SPREAD_10Y_3M); **change in the Euribor 3-month (3M_change);** change in the 10-year Government bond yield (10Y_change).

2. Foreign exchange

Nominal effective exchange rate, Euro area-17 countries vis-a-vis the EER-20 group of trading partners against Euro (FX).

3. Commodities

World market price of coal (COAL); change in Brent Crude (OIL_change); change in gold price (GOLD_change); change in copper price (COPPER_change); percentage change in Brent Crude (OIL_pchange); percentage change in gold price (GOLD_pchange); percentage change in copper price (COPPER_pchange).

4. Stock exchange

Change in EURO STOXX 50 Volatility Index VSTOXX (VSTOXX_change).

5. Corporate yields

Spreads between euro area Corporate AAA and BBB Merrill Lynch Bond Indexes for the following maturities: 1-3 years (AAA_BBB_1_3); 1-5 years (AAA_BBB_1_5); 1-10 years (AAA_BBB_1_10); 3-5 years (AAA_BBB_3_5); 5-7 years (AAA_BBB_5_7); 5-10 years (AAA_BBB_5_10); **7-10 years (AAA_BBB_7_10).** Spreads between euro area Corporate AAA Merrill Lynch Bond Indexes and Government Bond yields for the following maturities: 1-3 year corporate bond index vs 2-year government bond (AAA_1_3_2Y); 1-5 year corporate bond index vs 3-year government bond (AAA_1_5_3Y); 1-10 year corporate bond index vs 5-year government bond (AAA_1_10_5Y); 3-4 year corporate bond index vs 4-year government bond (AAA_3_5_4Y); 5-7 year corporate bond index vs 6-year government bond (AAA_5_7_6Y); 5-10 year

corporate bond index vs 8-year government bond (AAA_5_10_8Y); 7-10 year corporate bond index vs 10-year government bond (AAA_7_10_10Y); **7-10 year corporate bond index vs 6-year government bond (AAA_7_10_6Y)**.

B US financial data details

The following financial data have been used for estimating MIDAS models for the US. These series were chosen to correspond to the bolded series in the ECB technical appendix. The series is described, with the Haver DLX mnemonic given in parentheses.

1. Fixed income
LIBOR: Overnight (T111LON@INTDAILY); **LIBOR: Spot week** (T111L1W@INTDAILY); **LIBOR: 1 month** (T111L1M@INTDAILY); **LIBOR: 3 month** (T111L3M@INTDAILY); **LIBOR: 6 month** (T111L6M@INTDAILY); **LIBOR: 1 year** (T111L1@INTDAILY); **Government Bond Yield: 5 year** (T111G5@INTDAILY); **LIBOR 1 year and LIBOR 1 month spread**; **LIBOR: 3 month**, change.
2. Foreign exchange
Broad Trade-Weighted Exchange Rate (X111WTB@INTDAILY).
3. Commodities
Coal: level (P112GCB@INTDAILY); **Copper**: percentage change, level change (PZDCOP@DAILY); **Gold**: percentage change, level change (PZDGEI@DAILY); **Oil (Brent Crude)**: percentage change, level change (PZBRT@DAILY).
4. Stock exchange
CBOE Volatility Index: percentage change (SPVIX@DAILY).
5. Corporate yields
Corporate AAA and BBB Indexes: Spread for 15-year bonds (FMLCUA@DAILY); 15-year corporate bond index vs 5-year government bond.

Table 1: ECB Real-time GDP Early Estimates

Top panel: Schedule of real-time GDP Early Estimates - denoted EE - for quarter t on a bi-weekly schedule starting in month $t - 1$. Lower panel is the schedule against the real time data.

Quarter	t-1			t						t+1			
	1	2	3	1	2	3	1	2	3	1	2	3	
Month													
EE Release	1	2	3	4	5	6	7	8	9	10	11	12	
Errors	forecast			nowcast									backcast

Quarter	t-1			t						t+1			
	2	3	1	1	2	3	1	2	3	1	2	3	
Month													
EE Release	EE_t^1	EE_t^2	EE_t^3	EE_t^4	EE_t^5	EE_t^6	EE_t^7	EE_t^8	EE_t^9	EE_t^{10}	EE_t^{11}	EE_t^{12}	
Errors	$e_t^i, i = 1, \dots, 4$			$e_t^i, i = 5, \dots, 10$									$e_t^i, i = 11, 12$
Info	HF_t^1	HF_t^2	HF_t^3	HF_t^4	HF_t^5	HF_t^6	HF_t^7	HF_t^8	HF_t^9	HF_t^{10}	HF_t^{11}	HF_t^{12}	
	ΔGDP_{t-2}			ΔGDP_{t-1}						ΔGDP_{t-1}			
	$e_{t-2}^i, i = 1, \dots, 7$			$e_{t-1}^i, i = 8, \dots, 12$									

Table 2: One-Year ahead Forecasts of Real GDP growth - ECB and FRBNY

The table reports the mean forecast error (MFE), the mean squared forecast error (MSFE), the mean absolute forecast error (MAFE), root mean squared forecast error (RMSFE), minimum forecast error (MIN) and maximum forecast error (MAX). The entire period covers 2000-2012, the *Pre-Crisis* sample covers 2000-2007Q4 and the remainder of the sample corresponds to the crisis period. The ECB forecasts pertain to the fifth BMPE releases. The FRBNY are the last forecasts of each calendar year.

	MFE		MSFE		MAFE		RMSFE		MIN		MAX	
	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY
Full sample	0.96	1.67	8.83	9.82	1.85	2.04	2.97	3.13	-3.64	-3.23	13.27	11.84
Pre-Crisis	0.46	0.92	2.56	3.33	1.28	1.38	1.60	1.82	-2.20	-3.23	2.85	4.77
Crisis	1.64	2.40	17.41	16.15	2.62	2.70	4.17	4.02	-3.64	-2.27	13.27	11.84

Table 3: One-Year ahead Forecasts of Inflation - ECB and FRBNY

The table reports the mean forecast error (MFE), the mean squared forecast error (MSFE), the mean absolute forecast error (MAFE), root mean squared forecast error (RMSFE), minimum forecast error (MIN) and maximum forecast error (MAX). The entire period covers 2000-2012, the *Pre-Crisis* sample covers 2000-2007Q4 and the remainder of the sample corresponds to the crisis period. The ECB forecasts pertain to the fifth (B)MPE release - the December forecast prior to the target year. This particular release is picked as it is the best match with the end-of-year FOMC forecasts prior to the target year.

	MFE		MSFE		MAFE		RMSFE		MIN		MAX	
	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY	ECB	FRBNY
Full sample	-0.40	-0.15	0.94	0.30	0.79	0.45	0.97	0.55	-1.91	-1.00	2.75	1.12
Pre-Crisis	-0.49	-0.47	0.40	0.32	0.55	0.49	0.63	0.56	-1.40	-0.95	0.51	0.13
Crisis	-0.27	-0.05	1.69	0.30	1.11	0.44	1.30	0.55	-1.91	-1.00	2.75	1.12

Table 4: Real GDP growth forecasts of ECB across releases

The table reports the mean forecast error (MFE), the mean squared forecast error (MSFE), mean absolute forecast error (MAFE), root mean squared forecast error (RMSFE), minimum forecast error (MIN) and maximum forecast error (MAX). The table reports the mean forecast error (MFE), the mean squared forecast error (MSFE), mean absolute forecast error (MAFE), root mean squared forecast error (RMSFE), minimum forecast error (MIN) and maximum forecast error (MAX). The entire period covers 2000-2012, the *Pre-Crisis* sample covers 2000-2007Q4 and the remainder of the sample corresponds to the crisis period. The ECB forecasts pertain to the fifth BMPE releases. The FRBNY are the last forecasts of each calendar year.

Releases	Full sample			Pre-Crisis			Crisis		
	1	5	9	1	5	9	1	5	9
MFE	1.32	0.96	0.02	0.65	0.46	-0.29	2.09	1.64	0.50
MSFE	9.62	8.83	2.90	2.44	2.56	1.10	17.93	17.41	5.74
MAFE	1.84	1.85	1.11	1.26	1.28	0.86	2.51	2.62	1.49
RMSFE	3.10	2.97	1.70	1.56	1.60	1.05	4.23	4.17	2.40
MIN	-1.90	-3.64	-2.22	-1.90	-2.20	-2.22	-1.78	-3.64	-1.76
MAX	13.67	13.27	7.62	3.09	2.85	1.68	13.67	13.27	7.62

Table 5: One-Year ahead Forecasts FRBNY of Various Macroeconomic Series

The table reports the mean forecast error (MFE), the mean squared forecast error (MSFE), the mean absolute forecast error (MAFE), root mean squared forecast error (RMSFE). The entire period covers 2000-2012, the *Pre-Crisis* sample covers 2000-2007Q3 and the remainder of the sample corresponds to the crisis period. The series are Real GDP, PCE, BFI ETS, BFI Struc, Residential Investment, Exports, Imports, Government: Federal, Government: State and Local, PCE Deflator, and Core PCE Deflator.

	Real GDP	PCE	BFI ETS	BFI Struc	Residential Investment	Exports	Imports	Gov Federal	Gov S+L	PCE Deflator	Core PCE Deflator	
MFE												
Full sample	1.67	1.03	4.19	3.49	7.78	2.05	2.94	-0.21	2.48	-0.51	-0.15	
Pre-Crisis	0.92	0.81	5.43	-1.06	4.26	-2.29	2.21	-0.34	2.16	-1.16	-0.47	
Crisis	2.40	1.24	2.99	7.93	11.21	6.28	3.66	-0.09	2.79	0.13	-0.05	
MSFE												
Full sample	9.82	5.45	115.41	235.07	325.94	109.06	109.30	44.13	12.64	3.85	0.30	
Pre-Crisis	3.33	2.73	56.06	113.10	239.87	41.51	33.35	35.37	5.49	2.83	0.32	
Crisis	16.15	8.09	173.24	353.91	409.81	174.88	183.30	52.67	19.60	4.84	0.30	
MAFE												
Full sample	2.04	1.62	7.82	11.84	14.98	7.60	7.62	5.57	2.97	1.45	0.45	
Pre-Crisis	1.38	1.32	6.16	8.34	13.13	5.17	4.88	4.92	2.16	1.49	0.49	
Crisis	2.70	1.92	9.44	15.26	16.79	9.97	10.29	6.21	3.76	1.41	0.44	
RMSE												
Full sample	3.13	2.33	10.74	15.33	18.05	10.44	10.45	6.64	3.56	1.96	0.55	
Pre-Crisis	1.82	1.65	7.49	10.63	15.49	6.44	5.77	5.95	2.34	1.68	0.56	
Crisis	4.02	2.85	13.16	18.81	20.24	13.22	13.54	7.26	4.43	2.20	0.55	

Table 6: MIDAS Regression Analysis of Nowcasting Errors ECB Early Estimates Forecasts

The table reports the estimation results for MIDAS regressions (??) for the real-time GDP Early Estimates - or EE - of the ECB. The predictability of $e_t^i \equiv EE_t^i - \Delta GDP_t$, $i = 1, \dots, 12$, are reported. Provided are average R^2 values for the entire sample period, with MIDAS regressions using daily high frequency series detailed in Appendix ???. The sample runs from 2008Q2 to 2012Q4.

Releases	1	2	3	4	5	6
Ave R^2	0.31	0.25	0.34	0.33	0.38	0.39
MIDAS Reg.						
Ave R^2	0.04	0.04	0.03	0.07	0.03	0.03
AR(1)						
Bootstrap CI	[0.00, 0.24]	[0.00, 0.23]	[0.00, 0.24]	[0.00, 0.24]	[0.00, 0.24]	[0.00, 0.23]
Releases	7	8	9	10	11	12
Ave R^2	0.37	0.28	0.31	0.23	0.31	0.27
MIDAS Reg.						
Ave R^2	0.04	0.09	0.10	0.05	0.10	0.07
AR(1)						
Bootstrap CI	[0.00, 0.24]	[0.00, 0.25]	[0.00, 0.24]	[0.00, 0.23]	[0.00, 0.23]	[0.00, 0.24]

Table 7: MIDAS Regression Analysis of One-year GDP growth FRBNY Forecast Errors

Provided are average R^2 values for the entire sample obtained from MIDAS models using daily high frequency series detailed in Appendix ??, and compared with the R^2 for a simple AR(1) process for forecast errors. Also displayed is a 95% confidence interval for the AR R^2 . The dates are each FOMC meeting between early 2002 and the end of 2012, the *Pre-Crisis* sample covers 2000-2007Q4 and the remainder of the sample corresponds to the crisis period.

	Full Sample	Pre-Crisis	Crisis
Avg. R^2 MIDAS Reg.	0.31	0.19	0.37
Avg. R^2 AR(1)	0.02	0.03	0.03
Bootstrap AR(1) R^2 CI	[0.00, 0.08]	[0.00, 0.13]	[0.00, 0.14]

Figure 1: Euro area Real GDP projections by the Eurosystem/ECB staff

The figures report the published Eurosystem staff Broad Macroeconomic Projection Exercise (BMPE) and ECB staff Macroeconomic Projection Exercise (MPE) projections for Euro area real GDP and HICP in terms of ranges for the 2000-2013 sample. The solid horizontal lines across each year denote the actual outcomes. For each target year, nine forecast releases are published, as projections for each target year start to be published two years earlier in December and are last published in the December Monthly Bulletin of the target year. The midpoint of the range is also indicated for each of the nine prediction exercises. The solid horizontal lines denote the actual outcomes. (B)MPE projections for Euro area real GDP and HICP in terms of ranges. The solid horizontal lines denote the actual outcomes. For each target year, nine forecast releases are published, as projections for each target year start to be published two years earlier in December and are last published in the December Monthly Bulletin of the target year.

Figure 2: Euro area Harmonized Index of Consumer Prices projections by the Eurosystem/ECB staff

The figures report the published Eurosystem staff Broad Macroeconomic Projection Exercise (BMPE) and ECB staff Macroeconomic Projection Exercise (MPE) projections for Euro area the Harmonized Index of Consumer Prices (HICP) in terms of ranges for the 2000-2013 sample. The solid horizontal lines across each year denote the actual outcomes. For each target year, nine forecast releases are published, as projections for each target year start to be published two years earlier in December and are last published in the December Monthly Bulletin of the target year. The midpoint of the range is also indicated for each of the nine prediction exercises. The solid horizontal lines denote the actual outcomes.(B)MPE projections for Euro area HICP in terms of ranges. The solid horizontal lines denote the actual outcomes. For each target year, nine forecast releases are published, as projections for each target year start to be published two years earlier in December and are last published in the December Monthly Bulletin of the target year.

Figure 3: Probability of Real GDP Growth: Federal Reserve Bank of New York Forecasts against Survey of Professional Forecasters - April 2008

Figure 4: Housing Starts: Vintage Federal Reserve Bank of New York Forecasts During Great Recession

Figure 5: Real GDP growth and Core PCE Inflation under Alternative Scenarios

Figure 6: Fan Charts Real GDP growth and Core PCE Inflation

Figure 7: Scenario Probabilities and their Changes

Figure 8: Nominal FFR Scenarios

Figure 9: Historical Event-based Calibration: Depth of Recessions

Figure 10: Historical Event-based Calibration: Recovery Probabilities

Figure 11: Q-Q Plots - FRBNY Density Forecasts versus Survey of Professional Forecasts

The plots display Q-Q plots of the cross-section of SPF GDP growth forecasts against the FRBNY density forecasts produced in October of 2007-2011. The methodology to compute the density forecasts is described in section ??.

(a)(b)(c)
October-
to-to-to-
bebeber
2007-2011

(d)(e)(f)
October-
to-to-to-
bebeber
2010-2012