Els Opuscles del CREI

num. 37

September 2014

Density forecasts in economics and policymaking

Barbara Rossi

CREi

Centre de Recerca en Economia Internacional

The Centre de Recerca en Economia Internacional (CREI) is a research centre sponsored by the Universitat Pompeu Fabra (UPF) and the Generalitat de Catalunya. It is located at the campus of the UPF in Barcelona.

CREI's objective is to promote research in international economics and macroeconomics subject to the highest standards of academic scholarship.

Els *Opuscles del CREI* are the instruments for the diffusion of CREI's research beyond the academic world. Each *Opuscle*, aimed at a general public, gathers together conclusions and observations of specialized papers published, or about to be published, in professional journals. The opinions stated in *Els Opuscles del CREI* are the sole responsibility of the authors.

Pdf version of this *opuscle* and all the others *Opuscles del CREI* can be downloaded at: www.crei.cat/opuscles.php

Published by: CREI
Universitat Pompeu Fabra
Ramon Trias Fargas, 25-27 08005 Barcelona
Tel. 93 542 28 26

Design: Fons Gràfic
Printed by: Masanas Gràfiques

ISSN: 1137 - 7828
Legal diposit: DL B 20175-2014

Density forecasts in economics and policymaking*

Barbara Rossi

1. Forecasting: An introduction

1.1. Why forecasting?

How can we predict future outcomes in economics? How can we measure how uncertain future economic outcomes are? How can we determine whether our ignorance of the future is appropriately measured?

Predicting and evaluating future economic outcomes is crucial for making appropriate plans and assisting in the design and implementation of economic policies. Forecasts are produced, studied and evaluated everyday by central banks, academia, researchers in policy institutions, consumers, firms and practitioners. Central banks base their decisions about monetary policy on the analysis of the most likely future paths of a series of key macroeconomic variables, including inflation, output, and exchange rates, among others. Research and policy institutions, such as the International Monetary Fund and the World Bank, make recommendations based on the current as well as the predicted paths of key macroeconomic variables. Consumers plan their spending and saving decisions based on the expected, forecasted path of future interest rates and their income. Firms decide their prices and strategies based on expected, forecasted sales, and adjust their inventories based on the future costs of prime materials. Financial firms trade on the basis of their forecasts of asset values. Exporters and importers decide their purchases/sales based on the current as well as the future predicted value of the exchange rate. Several central banks (such as the European Central Bank, the Federal Reserve Bank of St. Louis, the Federal Reserve Bank of Philadelphia and the International Monetary Fund) maintain databases of surveys of professional forecasters that are routinely used to improve their own forecasts of future macroeconomic variables.

However, forecasting the future is not easy. The objective of this *opuscle* is to provide a review of density forecasts, with an emphasis on how they can be useful for policymakers and economists. We will first review some basic concepts in forecasting by discussing how predictions are typically made in economics and how one can evaluate whether they are appropriate. Then, we will discuss how density forecasts are different from traditional (point) forecasts and how density forecasts can be constructed and evaluated, which is the main goal of this *opuscle*. A practical example of forecasting U.S. real gross domestic product (GDP) is discussed to illustrate the methodologies and draw conclusions about whether output growth is predictable.

The empirical example is based, for simplicity of exposition, on a reduced-form autoregressive model. However, forecast densities can similarly be obtained without a model (such as survey forecasts, where individuals provide density forecasts based on their own judgment; e.g. the real-time

density forecast database collected by the Federal Reserve Bank of Philadelphia) or with a structural model (e.g. a Dynamic Stochastic General Equilibrium (DSGE) model; for example, see Rossi and Sekhposyan, 2014b). It is important to note that, however, the methods reviewed in this *opuscle* can be used no matter whether the density forecasts are obtained from a reduced-form or a structural model, or are survey based.

1.2. How are forecasts traditionally implemented? An example

Consider the case of a central bank interested in forecasting future real GDP growth. The availability of reliable output forecasts is very important for judging where the economy is heading to, for its implications on inflation and, consequently, for the monetary policy decisions that the Central Bank will make.

Model-based forecasts are typically obtained using a reference economic model. For example, the central bank staff periodically collects a series of macroeconomic variables, or predictors, which are believed to have been related historically to future output growth. Such variables are called in jargon "leading indicators". Then, the staff will measure the relationship between the leading indicators and output growth as accurately as possible by estimating either a structural model or a statistical model. The estimated model will then be used to produce forecasts of future output growth.

Forecasts can be reported in several ways. For example, they can be "point forecasts", e.g. forecasts of the expected value of the variable of interest (or target variable) in the future, like the ones we consider here; or forecasts of all the values that the target variable can take with a measure of their likelihood, that is, "density forecasts". The relationship between point and density forecasts is that the former is the mean of the density forecast; density forecasts more generally provide information on all the forecast quantiles. For example, density forecasts can be used to provide a forecast confidence interval, that is, an interval that should contain the future value with a pre-specified probability.

In what follows, we will first review how point forecasts are obtained and evaluated in practice, in order to define the terminology. The next section will describe how density forecasts are obtained and evaluated.

A leading example of a statistical model that we will use for illustration purposes is the autoregressive (AR) model. Let output growth at time "*i*" be denoted by y_i. The AR model is as follows:

$$\text{(AR)} \qquad y_i = \alpha + \beta\, y_{i-1} + e_i\,, \quad i=1,2,\ldots,t,$$

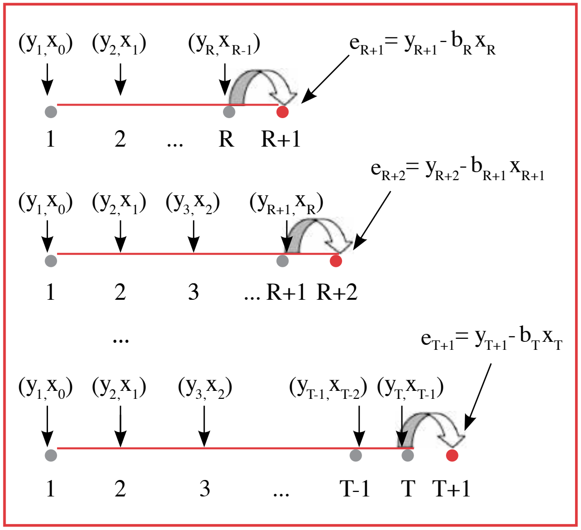
where e_i is an error term that measures the discrepancy between the model and the actual data. While output growth is observed, the parameters of the relationship (α and β) are unknown: at time t, the researcher will estimate them based on current and lagged observations of output growth.

Let the estimates of α, β obtained at time t be denoted, respectively, by a_t, b_t. Let the estimates be obtained using the full sample of data available up to the time the forecast is made, i.e. including observations *1* to *t*. Then, as time goes by, the parameter is re-estimated recursively over time. The forecasted value of y_{t+1} based on the available information at time t will simply be obtained as $f_{t+1|t} = a_t + b_t\, y_t$. The forecasts are then generated as time goes by, for $t=R,\ldots,T$.

Figure 1(a) illustrates how the forecasts are obtained in practice in a pseudo out-of-sample forecasting exercise. The latter exercise allows the researcher to obtain a dataset of out-of-sample point forecasts that can afterwards be evaluated by comparing them with the realized value of the forecasted variable. We will assume that the researcher starts his/her forecast procedure at time R and that he/she has a total sample of T+1 observations.[1] The researcher splits the total sample in two parts: the first part, from observations 1 to R ($R<T+1$) is used as the first estimation sample; the second part of P observations, including observations R+1 to T+1, is used to obtain and evaluate the forecasts. Thus, we mimic a situation where the researcher started his/her forecasting procedure at time R and attempt to replicate what he/she would have done in real-time as new data came in. At time R, he/she will estimate the model based on observations 1 to R, then produce the one period-ahead forecast for time R+1, $f_{R+1|R} = a_R + b_R\, y_R$. Then, when time R+1 comes, he/she will update his/her parameter values by re-estimating them using observations 1 to R+1, and produce the output growth forecast for time R+2. And so on and so forth, until the researcher arrives at time T, at which time he/she will re-estimate the parameters based on the sample of observations 1 to T, and produce the forecast for time T+1. This estimation procedure, where the parameters are recursively re-estimated over time using all the observations available until the time of the forecast, is called "recursive".

Often, however, the researcher realizes that the parameters might be changing over time, and, because of that, he/she feels that it is important to give more weight to the most recent observations. One simple way of achieving this is to use only the R most recent observations to estimate the parameters. In this case, the estimation procedure is the same as before at time R. However, when time R+1 comes, the researcher will update his/her parameter values by re-estimating them using observations 2 to R+1, and use those to produce the output growth forecast for time R+2. And so on and so forth, until the researcher arrives at time

Figure 1(a). **Recursive estimation scheme**



T, at which time he/she will re-estimate the parameters based on the sample of observations T-R+1 to T to produce the forecast for time T+1. Such forecasts are referred to as "rolling forecasts" (see e.g. West, 1996).

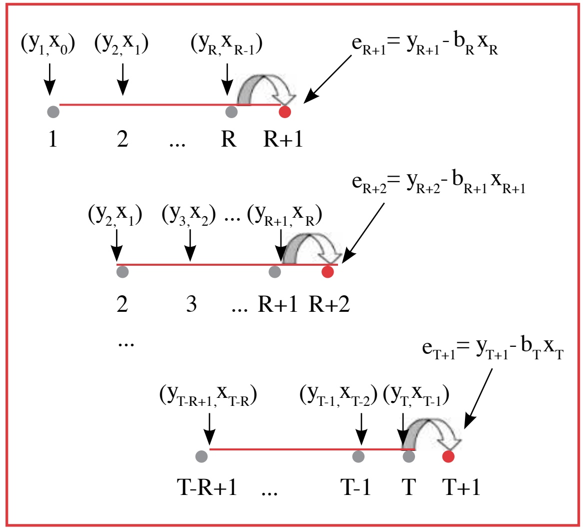
Figure 1 visualizes the differences between recursive and rolling forecasts. Figure 1(a) depicts the recursive estimation scheme while Figure 1(b) depicts the rolling one.

Models can be more complicated than the AR model, or can be structural (e.g. DSGE models), but the forecasts can be constructed in a similar way. For example, note that the autoregressive model we focused on can be easily extended to include other economic predictors as follows:

$$y_i = \alpha + \beta y_{i-1} + \zeta S_{i-1} + e_i\ ,\ i=1,2,\ldots,t$$

where S_{i-1} is the value of an additional predictor at time (i-1).

Figure 1(b). **Rolling estimation scheme**



Forecasts can also be constructed for longer horizons. Such multi-period ahead forecasts can be easily obtained in the simple AR model as follows. The researcher estimates the regression:

$$y_i = \alpha + \beta\, y_{i-h} + e_i\ ,\ i=1,2,\ldots,t$$

where, again, the estimates of α, β at time t are denoted, respectively, by a_t, b_t. In this case, $t=R$, $R+1,\ldots,T+h$. The forecasted value of y_{t+h} based on the available information at time t will simply be obtained as $f_{t+h|t} = a_t + b_t\, y_t$.[2]

1.3. How are forecasts traditionally evaluated?

Broadly speaking, there are two ways to evaluate forecasts. The first is an "absolute" way: we consider forecasts from a model and evaluate whether they satisfy certain "desirable" properties. For example, a desirable feature of a forecast is unbiasedness, that is whether on average the forecasts are close to the ex-post realized values of the target variable. A typical way to assess whether forecasts are unbiased is by evaluating whether the forecast error is zero on average. If that is the case, then there might be times at which the forecast over-predicts its target value and times at which the forecast under-predicts its target value; however, on average, the model does a good job, as over-predictions cancel out under-predictions.

Let the one-step ahead forecast error of the model be $e_{t+1|t} = y_{t+1} - f_{t+1|t}$. Typically, tests for forecast unbiasedness are implemented by regressing the forecast error on a constant, in order to test whether the forecast error has mean zero. That is, more formally, the researcher estimates the following regression:

$$e_{t+1|t} = \theta_1 + u_{t+1,t}$$

where $u_{t+1,t}$ is the error in the regression. Then, based on a t-test on θ_1, the researcher determines whether the forecast errors are zero on average. If the test rejects that θ_1 equals zero, then the forecasts are biased.

Another desirable property of forecast errors is that they should not be predictable based on information available at the time the forecast is made. In fact, if that is the case, then the researcher should have included the additional information in the model in order to improve its forecasts. Tests that are designed to evaluate whether that is the case are referred to as "tests for forecast rationality". They are implemented as follows. Let z_t be a variable omitted from model "AR", which the researcher suspects could be a useful predictor. Then the researcher estimates the following regression:

$$e_{t+1|t} = \theta_1 + \theta_2 z_t + u_{t+1,t}$$

where, as before, $u_{t+1,t}$ is the error in the regression. Then, based on a joint significance test on θ_1 and θ_2, the researcher determines whether the forecast errors are zero on average as well as whether they are correlated with the extra predictor. If the test cannot reject that θ_1 and θ_2 equal zero, then the forecasts are rational.[3]

An important point made in the literature is that, since the forecast error is estimated, rather than observed, researchers have to be careful when implementing their tests. In particular, they should correct the estimate of the standard errors to take into account parameter estimation error. West and McCracken (1998) discuss in detail how such correction works.

The second way to evaluate forecasts is by making comparisons among forecasting models. This evaluates models' "relative" forecasting ability. In this case, the researcher focuses on two or more models, rather than one. The objective is to evaluate whether the forecasting ability of the competing models, determined on the basis of a loss function selected by the researcher, is similar. A loss function typically used in the literature is the Mean Squared Forecast Error (MSFE), where the researcher compares the models based on their relative average squared values of the forecast errors.

The MSFE of a model is defined as follows:

$$\text{MSFE} = P^{-1} \sum e^2_{t+1|t}$$

where the summation ($\sum$) is intended to be from $t=R$ to T. That is, the researcher calculates the difference in the MSFEs of the two models and evaluates whether that is zero: if that is the case, the models' forecasting abilities are similar.

To evaluate whether the forecasting abilities of the two models are similar, the researcher tests whether the MSFE difference is zero (in expectation) by a t-test. Such a test is often referred to as the Diebold and Mariano (1995) and West (1996) test. Again, it is important to correct for parameter estimation uncertainty, as discussed in West (1996). See Diebold (2014) for a recent discussion of the use of these tests.

One important caveat to this procedure is the fact that comparing models according to their relative MSFEs may be problematic when the models under comparison are nested.[4] Clark and McCracken (2001), among others, offer alternative ways to compare nested models' forecasts.

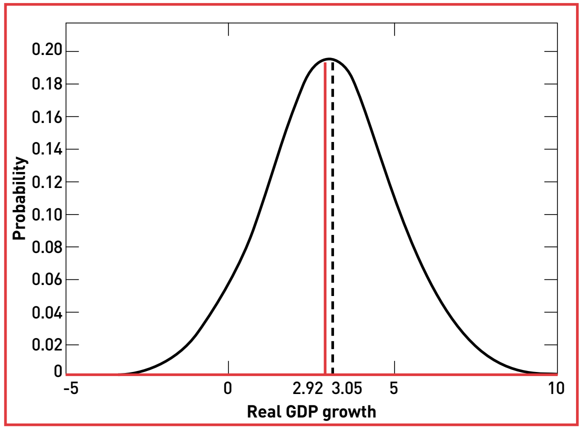
2. Density forecasts

2.1. What are density forecasts?

Central banks and policy institutions have recently started to realize that it is important to estimate and report the uncertainty around their forecasts. In fact, the forecast defined previously (point forecast) measures the central tendency of the target variable y, or the best forecast; however, because this is an estimate, there is uncertainty around it. Quantifying this uncertainty is important to convey how "sure" the researcher is regarding the precision of the forecasted value.

One way to report the uncertainty around point forecasts is to use density forecasts. Density forecasts summarize the information regarding the estimated forecast distribution. For example, in the simple AR model previously described, the researcher will make an assumption on the error term, e_i, and derive the density forecast based on that. A typical assumption is that the error term is

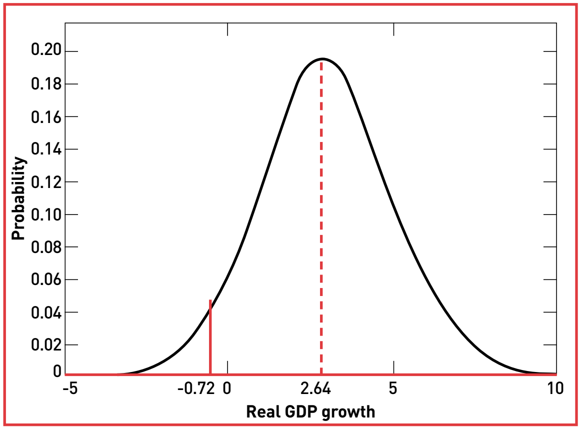
Figure 2(a). One quarter-ahead density forecast of annualized U.S. real GDP growth made in 2004Q2 (Based on normality)



normally distributed. It follows that, by construction, the error term should be mean zero (if the forecast is unbiased). The researcher will typically proxy the unknown variance of the forecast error with the estimated variance of the in-sample fitted errors. The in-sample fitted errors can be easily obtained as: y_i-a_t-$b_t\,y_{i-1}$, for i=$1,2,\ldots,t$. Finally, the density forecast can be obtained as follows. Conditional on information at time t, the forecast $f_{t+1|t}$ is normally distributed with mean $a_t + b_t y_t$. Its variance is proxied by the variance of the in-sample fitted errors.

To visualize and clarify what a density forecast is, Figure 2(a) plots a forecast density made in 2004Q2 for the following quarter, using the autoregressive model, eq. (AR). The x-axis reports the possible values of real GDP growth one quarter into the future,[5] and the y-axis reports the probability that such value will realize. According to the model, the mean growth rate of real GDP should have been 3.05 in the third quarter of 2004, as the dashed bar shows. The density forecast is centered

Figure 2(b). One quarter-ahead density forecast annualized U.S. real GDP growth made in 2008Q1 (Based on normality)

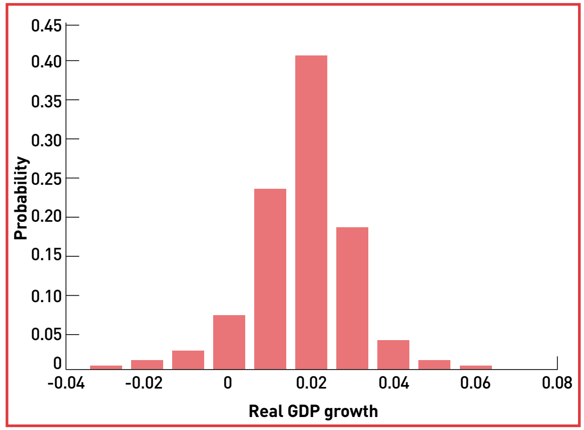


on that value and normally distributed around it. The realization was instead 2.92, indicated by a solid bar in the figure. Note how close the actual realization is to the forecasted value.

Figure 2(b) plots instead a forecast density made in 2008Q1 for the following quarter, using the same model.[6] According to the model, the mean growth rate of real GDP should have been 2.64 in the second quarter of 2008. Again, the density forecast is centered on that value and normally distributed around it. Note the high uncertainty around the mean forecast. The realization was instead -0.72, indicated by a bar in the figure. Clearly, the actual realization was considered ex-ante a very unlikely event. In fact, one can calculate the ex-ante probability of observing a growth rate less than or equal to -0.72 based on the density forecast in the figure, and that was only 0.04.

While, typically, density forecasts are obtained under an assumption on the distribution of the error term, this is not always the case. Another

Figure 2(c). One year-ahead density forecast annualized U.S. real GDP growth made in 2008Q1 (From the Survey of Professional Forecasters)



example is density forecasts obtained from survey forecasts. In that case, the survey will directly provide the forecast distribution. Such quantiles are obtained by directly asking the survey respondents what is the probability that the target variable will be in certain ranges. For example, the survey will ask the respondents what is the probability that inflation will be between 0 and 1 percent a year from now; what is the probability that it will be between 1 and 2 percent; and so forth. Survey forecasts will typically report the forecast distribution of each respondent and/or the average across respondents.

Figure 2(c) reports Survey of Professional Forecasters (SPF) density forecasts made in 2008Q1 for 2009. Data are from the Federal Reserve Bank of Philadelphia. Since the dataset collects forecasts for the coming year (2009), they are not directly comparable with Figures 2(a-b).[7] Nevertheless, our goal is not to compare different forecast densities, but just to illustrate how they can be constructed using different methodologies. Figure 2(c) shows that the SPF density forecast puts a high probability (0.40) on a growth rate around 0.02. Again, the realized value is very different, and equals -0.02, such an unlikely event according to the forecasters, since they predicted it to happen with a probability equal to 0.01.

2.2. How are density forecasts currently used in economics, forecasting and policymaking?

A natural application of forecast densities is the analysis and evaluation of macroeconomic risk. For example, before every Federal Open Market Committee meeting, the Federal Reserve Bank of New York complements its best guess of the future path of key macroeconomic variables (the so-called "modal forecast") with an assessment of the risk around it (Alessi et al., 2014). The uncertainty is evaluated based on the most prominent risk scenarios using judgment, and by a robustness analysis of the conditioning assumptions under which the modal forecast was obtained.

Simulations of alternative scenario-driven paths also generate forecast distributions that can be used to evaluate how precise the modal forecasts are and whether the risk of overestimating the outcome is higher or smaller than the risk of underestimating it.

Typically, at central banks, density forecasts are produced conditional upon a projected scenario for several key variables, one of which is the interest rate set by the central bank itself. The idea is to analyze how forecasts change depending on the projected path of the instrument of monetary policy, i.e. the interest rate. By comparing forecasts as well as density forecasts associated with several scenarios, the policymaker will evaluate the effects of alternative monetary policy choices. The comparison will provide important information which will ultimately guide monetary policy decisions.

Figure 3(a). One quarter-ahead forecasts of annualized U.S. real GDP growth, their uncertainty and actual realizations

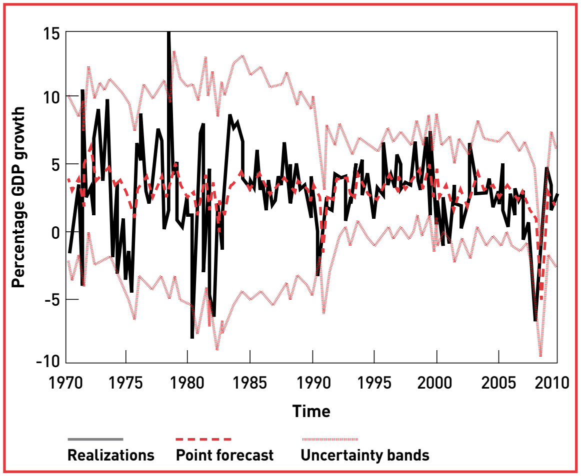


Figure 3(a) visualizes one quarter-ahead forecasts of annualized U.S. real GDP growth, together with the uncertainty surrounding it and the actual realization. The model is the AR model previously introduced. The forecasts are implemented with a rolling estimation scheme using a window size of 10 years. The continuous line plots the realized GDP growth, the dashed line plots its forecast one-quarter in advance, and the outer bands plot the uncertainty around the forecasts (interpreted as a 95% confidence band based on rolling estimates of the variance of the in-sample fitted errors).

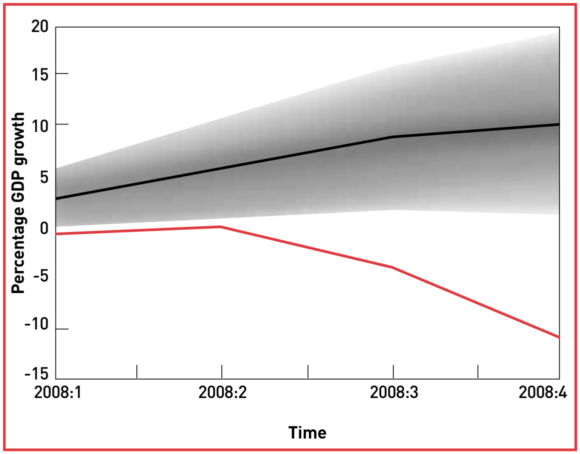
Several interesting observations emerge from the picture. First, GDP growth volatility was much higher prior to 1984. In fact, due to the decrease in volatility observed after 1984, the time period starting in 1984 and lasting until the financial crisis of 2007–2008 was referred to as the "Great Moderation". Note that the measure of uncertainty around the forecast is obtained using a rolling window of 10 years. While this measure can be obtained in real-time, it is nevertheless slow to adapt to the changing environment. The figure also clearly reveals the financial crisis of 2007–2008, marked by a substantial decrease in real output growth. Note how the forecast failed to predict the large drop in GDP caused by the crisis, an observation consistent with the analysis in Figure 2(b). The realization is not even included within the uncertainty measure, even though the measure is quite wide. This implies that the large magnitude of the drop in output growth during the financial crisis was essentially unpredictable using the AR model, and that the actual realization was considered an event that might have happened with less than 5% probability. Clearly, however, the forecasts are fast in catching up the drop in output growth and its subsequent improvement with a one-quarter delay. On the other hand, note that the recession in 1984 (as well as the recessions caused by oil price shocks in the mid-1970s) were within the uncertainty measure. Finally, note that the forecast lags the actual data; this is a typical feature of models based on lagged data, as the one we consider here, as they have difficulties in predicting turning points.

Note that the uncertainty plotted in the figure is such that the bands should include the realization with 95% probability. Therefore, they correspond to the 2.5-th and 97.5-th quantiles of the distribution. One could have plotted other quantiles of the distribution, or several quantiles in the same picture. However, the latter may result is a picture that is difficult to read. Instead, to convey information on several forecast quantiles at the same time, researchers typically plot fan charts.

The use of density forecasts and especially fan charts in central banks was pioneered by the Bank of England. To ease the communication of the target inflation rate to the public, the Bank of England decided to report not only the point forecast for inflation, but also the range of values of inflation that they deem most likely. For example, they report in shaded areas the values of inflation that are expected with probability, say, 90%; they also report the values that are expected with probability 80%, 70% and so forth until 10%. The values of expected inflation corresponding to the margins of these areas are the deciles of the forecast density distribution. The way they present this information is by depicting the areas in progressively lighter colors, the darkest colors being associated with the area that represents the most likely outcome (the values of inflation that are expected with probability 90%) and the lightest color being associated with the area that represents the most uncertain outcome (the values expected with probability 10%). As the forecast density can be obtained for several periods ahead, it is common to report these areas as a function of the forecast horizon. Typically, as the forecast horizon increases, the uncertainty increases as well (as events further in the future are more difficult to predict accurately), thus the areas become wider, with the plot resembling a fan; hence the name "fan chart".

Figure 3(b) depicts a fan chart for forecasts made in 2007:Q4 for the four quarters of 2008. The continuous red line depicts the realized cumulative rate of growth since 2008:Q1; the shaded area depicts the deciles of the forecast distribution around the mean forecast, depicted as a solid line. The 90^{th} decile is the lightest area, and the 10^{th} decile is the darkest area. Clearly, the simple autoregressive model predicted positive growth before the financial crisis, and the realizations are quite outside the predicted quantiles.

Figure 3(b). **Fan chart of annualized, cumulative U.S. real GDP growth**



3. Evaluating density forecasts

Since density forecasts play such an important role in providing information on the uncertainty around point forecasts, it is crucial to evaluate whether they are well specified. If density forecasts are not correctly specified, then the measure of uncertainty that they provide is incorrect.

In particular, density forecasts can be evaluated according to the same two broad categories discussed above, that is, they can be evaluated in terms of their "absolute" or in terms of their "relative" predictive performance. In this *opuscle*, we will focus on the former.

There are several ways to evaluate the correct specification of density forecasts. A typical approach is to use probability integral transforms.[8] See Corradi and Swanson (2006b) for a detailed overview.

3.1. Description of methodologies and their implementation

The use of probability integral transforms to evaluate density forecasts was pioneered by Diebold, Gunther and Tay (1998) and Diebold, Tay and Wallis (1999). A probability integral transform (PIT) is the cumulative probability evaluated at the actual, realized value of the target variable. It measures the likelihood of observing a value less than the actual realized value, where the probability is measured by the density forecast.

One important finding in the literature dates back to Diebold et al. (1998). They demonstrate that the PIT is uniform, independent and identically distributed if the density forecast is correctly specified. Therefore, Diebold et al. (1998) propose to test the correct specification of density forecasts by testing whether the PIT is uniformly distributed and independent.

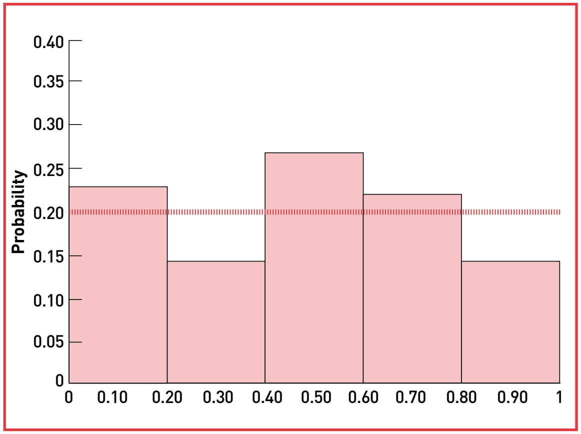
3.2 Uniformity

The uniformity property means that the probability that the realized value is higher (lower) than the forecasted value is the same (on average over time) no matter whether we consider high realizations or low realizations of the variable we are forecasting.

The test for uniformity may involve plotting the empirical distribution of the PIT (or histogram). Uniform data would have an empirical distribution function that looks like a rectangle. The further the empirical distribution of the PITs is from a rectangle, the stronger the evidence against correct specification of the density forecast.

Figure 4 reports the empirical distribution of the PIT associated with one quarter-ahead density forecasts from the AR model. The distribution

Figure 4. Empirical distribution of the PIT for one quarter-ahead density forecats of annualized U.S. real GDP growth using the AR model (Based on normality)



resembles that of a uniform distribution[9] and suggests that the density is correctly specified.

One could also implement more sophisticated tests. For example, Diebold et al. (1998) consider applying traditional tests for uniformity too (such as Cramer-von Mises and Kolmogorov-type tests).

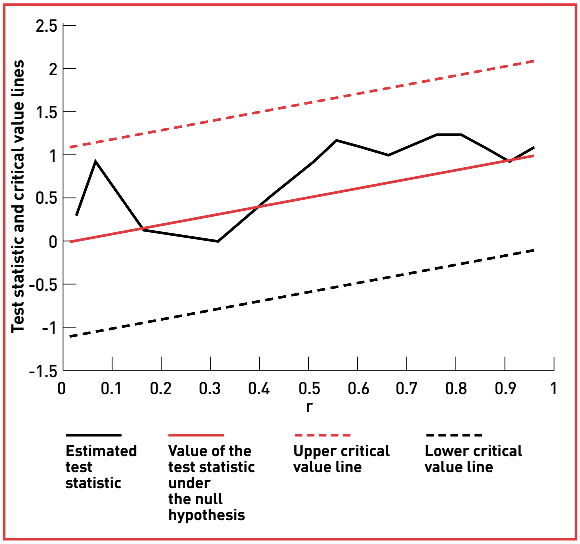
How different is the empirical distribution function from the theoretical (uniform) distribution? The empirical distribution function of the PIT is estimated from the data, and therefore one should take estimation error into account when comparing the empirical and the uniform distribution. Approaches to test for uniformity include the test proposed by Corradi and Swanson (2006a) and Rossi and Sekhposyan (2014b). They both have the advantage of taking into account parameter estimation error. The similarity in the two papers is that they both focus on the empirical cumulative distribution function of the PIT. If the PITs were uniform, their empirical cumulative distribution function would resemble a 45-degree line. The difference between Corradi and Swanson (2006a) and Rossi and Sekhposyan (2014b) is the way they handle parameter estimation uncertainty: the former allow for a large estimation window size whereas the latter assume a fixed estimation window size. The former is more appropriate when the researcher aims at evaluating forecast densities of models as if their parameters were precisely estimated in the sample.[10] In contrast, the latter is more appropriate in situations where researchers realize that they may not be able to precisely estimate the parameter value in finite sample and they want to evaluate whether the forecast density is correctly specified at the actual, estimated parameter value (rather than at the population value).

When the Rossi and Sekhposyan (2014b) test is applied to the data, it does not reject the correct specification of forecast densities from the AR model based on a normal distribution. Figure 5 shows the results in detail. The figure plots the cumulative empirical distribution function estimated from the data (dark-black solid line). The figure also plots the 45-degree line, which represents the cumulative distribution function of a uniform distribution (light-red solid line), together with the confidence bands based on Rossi and Sekhposyan's (2014b) statistic (dashed lines). As the empirical distribution function is within the confidence bands, one concludes that the normal distribution provides a good approximation to the forecast density.[11]

3.3. Independence

Empirical evidence in favor of uniformity of the PITs means that, on average, the unconditional distribution is correctly specified. Even if uniformity is not rejected, the pattern of rejection may be non-random over time, which raises concerns

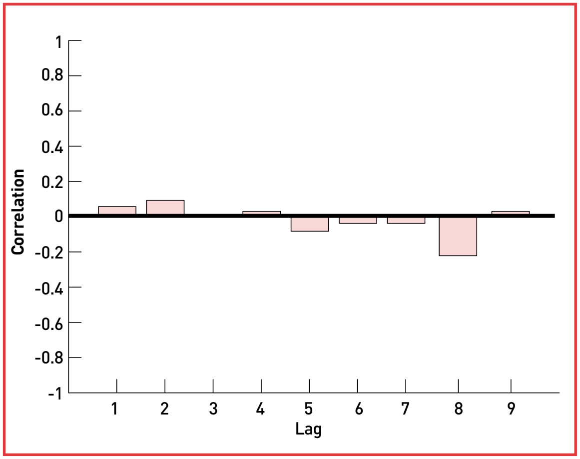
Figure 5. Empirical cumulative distribution function of the PITs for one quarter-ahead forecasts of annualized U.S. GDP growth forecasts using the AR model (Based on normality) and bands for the test of correct specification



about lack of independence of the PITs. Lack of independence in the PITs, thus, signals misspecification of the density forecast.

Diebold et al. (1998) propose to test for independence of the PITs by using tests for uncorrelatedness.[12] Figure 6 plots the empirical correlation function of the PITs, or correlogram. The figure plots the correlation between the PIT for forecasting output growth and its lagged value for a given lag length; the lag length is reported on the x-axis, and the correlation on the y-axis. If, at all lags bigger than or equal to one, the correlation is close to zero, this implies that the PITs are serially uncorrelated and it is interpreted as empirical evidence in favor of the correct specification of the density forecast.[13] The serial correlation at lag equal to one

Figure 6. Correlagram of PITs for one quarter-ahead forecasts of annualized U.S. GDP growth forecasts using the AR model (Based on normality)



equals 0.0689, at lag equal to two equals 0.1075, at lag three equals 0.0064, etc. Thus, it is pretty small which again suggests correct specification of the normal density forecast.

3.4. Identical distribution

The result by Diebold et al. (1998) implies that, under correct specification, the PIT is uniform, independent and identically distributed. However, Diebold et al. (1998) only propose testing the first two properties. Rossi and Sekhposyan (2013) propose to additionally test whether the PIT is identically distributed.

Rossi and Sekhposyan's (2013) test is useful for the following reason. All the tests previously considered are invalid unless the data are stable over time. However, macroeconomic data are subject to changes or instabilities. For example, it is well-known that the predictive ability of several macroeconomic variables for output growth and inflation has changed over time (see for example Stock and Watson, 1996, and Rossi, 2013) or that monetary policy has changed over time (see for example Clarida, Gali and Gertler, 2000). It might therefore be possible that the preferred model used by central banks might be misspecified in some sub-samples, but not in others. Such changes over time invalidate the traditional procedures. Rossi and Sekhposyan (2013) propose to test the correct specification in sub-samples of the data, and therefore their analysis can detect misspecification even when it appears only in a sub-sample of the data.[14]

3.5. Evaluating the "relative" performance of density forecast

One way to evaluate whether two competing models' density forecasts are sufficiently similar (or, whether they are similarly distant from the true distribution of the target variable) is that proposed by Amisano and Giacomini (2007) and Diks, Panchenkob and van Dijk (2011). Amisano and Giacomini's (2007) procedure is similar in spirit to the tests of relative forecasting performance except that, rather than comparing the MSFEs of the models, they propose to compare (the log of) the predictive densities of the two models. In particular, they propose a t-test on the difference of (the log of) the predictive densities.[15]

4. Conclusions

This *opuscle* has provided a review of density forecasts. First, we have provided a detailed description of how to construct forecasts, in particular we have reviewed recursive and rolling estimation schemes. Then, we have provided a detailed description of how density forecasts can be produced, and their merits relative to traditional (point) forecasts. Finally, we have reviewed tests to evaluate how 'good' point and density forecasts are, which can help researchers assess the correct specification in their analyses.

Throughout, we have provided an illustration to point and density forecasts of U.S. real GDP growth obtained via an autoregressive model and normality assumptions. We have shown how difficult it is to forecast real output growth during special circumstances such as a financial crisis, and how a simple AR model with normal disturbances seems to provide a good forecasting environment.

While we have illustrated the concepts using an example based on forecasting real GDP growth using an AR model with a normal distribution, the analysis can be extended to other models, other macroeconomic target variables and other types of distributions. Rossi and Sekhposyan (2014a) evaluate normal conditional predictive densities for both U.S. output growth and inflation using several well-known forecasting models that rely on a large number of macroeconomic predictors. They find that normality is rejected for most models they consider (autoregressive distributed lag, factor models and Bayesian VAR models), at least in some dimensions. Interestingly, however, they find that combinations of predictive densities are correctly approximated by a normal density.[16]

An important avenue for future research is to improve the early detection of turning points, i.e. special situations where the behavior of the data suddenly changes. In particular, analyzing whether and how the tools available in the literature can provide not only detection of misspecification in point and density forecasts, but also guidance on how to improve the forecasting models themselves.

Notes

(*) Acknowledgments. I thank Tatevik Sekhposyan and Gergely Ganics for comments. Some of the research described in the present opuscle *has been funded by the European Commission Marie Curie Grant 303434.*

(1) The notation follows West (1996).

(2) The latter are referred to as "direct forecasts" because the parameters are estimated and used directly to produce the forecast, as opposed to "iterated forecasts" which are instead obtained by estimating eq. (AR) and iterating it forward to produce the h-step ahead forecast as: $f_{t+h|t} = a_t h + b_t^h y_t$. *Thus, direct forecasts can be obtained directly from the estimated model without transforming the estimated parameters, whereas iterated forecasts cannot. See Marcellino, Stock and Watson (2006) for a comparison between iterated and direct forecasts.*

(3) Note that this jointly evaluates whether the forecast errors are unbiased as well as whether the additional predictor may be useful for forecasting the target variable.

(4) Two models are nested when one can be obtained as a special case of the other, e.g. by imposing zero restrictions on the parameters. For example, the AR model, $y_i = \alpha + \beta y_{i-1} + e_i$, *is nested in the model* $y_i = \alpha + \beta y_{i-1} + \zeta S_{i-1} + e_i$.

(5) The data are from the Federal Reserve Economic Data of the St. Louis Fed (FRED). The mnemonics is rgdp@us.

(6) Note that 2008Q1 is a period associated with the recent financial crisis.

(7) This is so because, while they are made in 2008Q1, they are forecasting year-on-year output growth in 2009, that is, one year ahead. Thus the forecast horizon is different as well.

(8) Alternative approaches include log probability scores or other scoring rules.

(9) The distribution function of the uniform distribution is depicted by the dotted line.

(10) That is, the null hypothesis of correct specification of the forecast density holds at the (pseudo-true) population value of the parameter.

(11) The Rossi and Sekhposyan (2014b) sup-type test value is 0.82 for one quarter-ahead forecasts and 1.13 for four quarter-ahead forecasts. The Rossi and Sekhposyan (2014b) mean-type test value is 0.15 for one quarter-ahead forecasts and 0.25 for four quarter-ahead forecasts. None of these statistics are significant.

(12) Since independence implies uncorrelatedness, lack of uncorrelatedness can be interpreted as lack of independence. However, if the test does not reject, it does not necessarily mean that the data are independent, only that they are uncorrelated.

(13) More precisely, as previously mentioned, uncorrelatedness does not imply independence, but lack of uncorrelatedness implies lack of independence. So the correct interpretation is that the test does not provide enough empirical evidence against the correct specification of the density forecast.

(14) Alternative tests include Berkowitz (2001). Berkowitz (2001) proposes to focus on the inverse normal of the PIT. If the PIT is uniform, then its inverse normal is normal. Thus, Berkowitz (2001) proposes to estimate an autoregressive model for the inverse normal of the PIT and test whether the mean is zero, the variance is one, and the correlation is zero. The advantage of Berkowitz's (2001) approach is that its test can be implemented with a likelihood ratio test. On the other hand, Berkowitz's (2001) approach focuses on testing specific moments of the empirical distribution function. In other words, by estimating an autoregressive process with one lag for the PITs, the approach will only test the first two moments (the mean and the variance) and the first-order correlation but will not test higher moments. Diebold et al. (1998) focus instead on the whole distribution function, and therefore consider all the moments simultaneously. While Berkowitz's (2001) test could be implemented on higher moments (such as skewness, kurtosis, etc.), the higher the number of moments the researcher considers, the less precise the test will be in finite samples. When the Berkowitz (2001) test is applied to the AR model with normal disturbances considered here, it does reject the correct specification of the mean, but not that of the serial correlation, and the joint test (on the mean, variance and lack of serial correlation) does not reject the correct specification of the model at standard critical values.

(15) Their test is slightly more general because it allows the researcher to give more importance to certain regions of the density forecast than others. For example, a researcher may be more interested in comparing the performance of the models in the most unlikely events, i.e. in the tails of the distribution.

(16) Other empirical works focusing on the estimation and evaluation of density forecasts include Clark (2011), Garratt et al. (2003), Jore, Mitchell and Vahey. (2010), Manzan and Zerom (2009), Mitchell and Wallis (2011) and Clements and Smith (2000)

References

Alessi, L., E. Ghysels, L. Onorante, R. Peach and S. Potter (2014), "Central bank macroeconomic forecasting during the global financial crisis: The European Central Bank and the Federal Reserve Bank of New York experiences", *Journal of Business and Economic Statistics*, forthcoming.

Amisano, G. and R. Giacomini (2007), "Comparing density forecasts via weighted likelihood ratio tests", *Journal of Business and Economic Statistics*, 25(2), 177–190.

Berkowitz, J. (2001), "Testing density forecasts, with applications to risk management," *Journal of Business and Economic Statistics*, 19, 465–474.

Clarida, R., J. Gali, and M. Gertler (2000), "Monetary policy rules and macroeconomic stability: evidence and some theory", *Quarterly Journal of Economics*, 115, 147–180.

Clark, T. E. (2011), "Real-Time density forecasts from bayesian vector autoregressions with stochastic volatility", *Journal of Business and Economic Statistics*, 29(3), 327–341.

Clark, T. E. and M. W. McCracken (2001), "Tests of equal forecast accuracy and encompassing for nested models", *Journal of Econometrics*, 105, 85–110.

Clements, M. P. and J. Smith (2000), "Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment," *Journal of Forecasting*, 19, 255–276.

Corradi, V. and N. R. Swanson (2006a), "Bootstrap conditional distribution tests in the presence of dynamic misspecification", *Journal of Econometrics*, 133, 779–806.

Corradi, V. and N. R. Swanson (2006b), "Predictive density evaluation", in G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Elsevier, 197–284.

Diks, C., V. Panchenkob and D. van Dijk (2011), "Likelihood-based scoring rules for comparing density forecasts in tails", *Journal of Econometrics*, 163, 215–230.

Diebold, F. X. (2014), "Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold-Mariano tests", *Journal of Business and Economic Statistics*, forthcoming.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998), "Evaluating density forecasts with applications to financial risk management", *International Economic Review*, 39(4), 863–883.

Diebold, F. X. and R. Mariano (1995), "Comparing predictive accuracy", *Journal of Business Economics and Statistics*, 13, 253–263.

Diebold F. X., A. S. Tay and K. F. Wallis (1999), "Evaluating density forecasts of inflation: the survey of professional forecasters", in Engle R.F. and H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, 76–90.

Garratt, A., K. Lee, M. H. Pesaran and M. H. Shin (2003), "Forecast uncertainties in macroeconomic modeling; an application to the U.K. economy", *Journal of the American Statistical Association*, 98, 829–838.

Jore, A. S., J. Mitchell and S. P. Vahey (2010), "Combining forecast densities from VARs with uncertain instabilities," *Journal of Applied Econometrics*, 25, 621–634.

Manzan, S. and D. Zerom (2009), "Are macroeconomic variables useful for forecasting the distribution of U.S. inflation?," MPRA Paper 14387, University Library of Munich.

Marcellino, M., J. H. Stock and M. W. Watson (2006), "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series", *Journal of Econometrics*, 135, 499–526.

Mitchell, J. and K. Wallis. (2011), "Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness," *Journal of Applied Econometrics*, 26, 1023–1040.

Rossi, B. (2013), "Advances in forecasting under instabilities", in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2, Elsevier-North Holland Publications, 1203–1324.

Rossi, B. and T. Sekhposyan (2013), "Conditional predictive density evaluation in the presence of instabilities", *Journal of Econometrics*, 177(2), 199–212.

Rossi, B. and T. Sekhposyan (2014a), "Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set", *International Journal of Forecasting*, forthcoming.

Rossi, B. and T. Sekhposyan (2014b), "Alternative tests for correct specification of conditional forecast densities", *mimeo*.

Stock, J. H. and M. W. Watson (1996), "Evidence on structural stability in macro-economic time series relations," *Journal of Business and Economic Statistics*, 14, 11–30.

West, K. D. (1996), "Asymptotic inference about predictive ability," *Econometrica*, 64, 1067–1084.

West, K. D. and M. W. McCracken (1998), "Regression-based tests of predictive ability," *International Economic Review*, 39, 817–840.

Opuscles already published

1. **Reconsidering Spanish unemployment**
Ramon Marimon (June 97)
2. **Reducing unemployment. At any cost?**
Fabrizio Zilibotti (December 97)
3. **Capital and labor taxes, macroeconomic activity, and redistribution**
Albert Marcet (November 98)
4. **The lender of last resort in today's financial environment**
Xavier Freixas (November 99)
5. **Why does the public sector grow? The role of economic development, trade and democracy**
Carles Boix (November 99)
6. **Gerontocracy and social security**
Xavier Sala-i-Martin (July 2000)
7. **The political viability of labour market reform**
Gilles Saint-Paul (December 2000)
8. **Are EU policies fostering growth and reducing regional inequalities?**
Fabio Canova (May 2001)
9. **Agglomeration effects in Europe and the USA**
Antonio Ciccone (September 2001)
10. **Economic polarization in the Mediterranean basin**
Joan Esteban (May 2002)
11. **How do households invest their wealth?**
Miquel Faig (October 2002)
12. **Macroeconomic and distributional effects of social security**
Luisa Fuster (April 2003)
13. **Educating intuition: A challenge for the 21st century**
Robin M. Hogarth (September 2003)
14. **Capital controls in post-war Europe**
Hans-Joachim Voth (April 2004)
15. **Taxation of financial intermediaries**
Ramon Caminal (September 2004)
16. **Ready to take risks? Experimental evidence on risk aversion and attraction**
Antoni Bosch-Domènech / Joaquim Silvestre i Benach (November 2005)
17. **Social networks and labour market outcomes**
Antoni Calvó-Armengol (January 2006)
18. **The effects of employment protection in Europe and the USA**
Adriana D. Kugler (February 2007)
19. **Urban sprawl: Causes and consequences**
Diego Puga (January 2008)
20. **Western European long term growth, 1830-2000: Facts and issues**
Albert Carreras and Xavier Tafunell (June 2008)
21. **Overcoming coordination failure in firms and organizations: Experimental evidence**
Jordi Brandts (March 2009)
22. **The misallocation of talent**
José V. Rodríguez Mora (May 2009)
23. **Complementarities in innovation strategy and the link to science**
Bruno Cassiman (September 2009)
24. **Simple mechanisms to resolve conflicting interests and to share the gains**
David Pérez-Castrillo (November 2009)
25. **Transfer of university innovations**
Inés Macho-Stadler (January 2010)
26. **Firing costs, dismissal conflicts and labour market outcomes**
Maia Güell (June 2010)
27. **Inequality and tax progressivity**
Juan Carlos Conesa (October 2010)
28. **Happiness economics**
Ada Ferrer-i-Carbonell (May 2011)
29. **School choice in Spain: Theory and evidence**
Caterina Calsamiglia (September 2011)
30. **Corporate social responsibility and social welfare. How to promote socially responsible business strategies**
Juan-José Ganuza (March 2012)
31. **The effectiveness of foreign aid: From the macroeconomic perspective to the experimental evaluation**
José G. Montalvo / Marta Reynal-Querol (June 2012)
32. **Fiscal policy in the European Monetary Union**
Evi Pappa (September 2012)
33. **How important are capital markets Imperfections in determining firm decisions and aggregate fluctuations?**
Andrea Caggese (November 2012)
34. **Globalization, technology and inequality**
Gino Gancia (April 2013)
35. **Credit cycles and systemic risk**
José-Luis Peydró (December 2013)
36. **The labour market impact of immigration**
Albrecht Glitz (June 2014)
37. **Density forecasts in economics and policymaking**
Barbara Rossi (September 2014)

Barbara Rossi

Barbara Rossi is an ICREA Professor of Economics at Universitat Pompeu Fabra. She previously has been an Associate Professor with tenure at the department of Economics at Duke University, after earning her Ph.D. from Princeton University (2001). She is a CEPR Fellow, a member of the CEPR Business Cycle Dating Committee and a Director of the International Association of Applied Econometrics. She served as associate editor of the *Journal of Business and Economic Statistics*, the *Journal of Economic Dynamics and Control*, and the *Journal of Applied Econometrics*.

Professor Rossi specializes in the fields of time series econometrics, as well as applied international finance and macroeconomics. Her current research focuses on forecasting and macroeconometrics. Professor Rossi has published her research findings in the *Review of Economic Studies, Quarterly Journal of Economics*, the *Journal of Business and Economic Statistics*, the *International Economic Review, Econometric Theory*, the *Journal of Applied Econometrics*, the *Journal of Money, Credit and Banking, Journal of Econometrics*, the *Review of Economics and Statistics,* and *Macroeconomic Dynamics*. She has presented her findings at a variety of professional conferences and meetings, including the SED meetings, the Econometric Society Meetings, the Joint Statistical Meetings, the NBER-NSF Time Series Conference, the NBER, as well as the AEA meetings.

She received two National Science Foundation grants, and was recently invited to write a chapter on "Advances in Forecasting under Model Instabilities" for the *Handbook of Economic Forecasting* (Elsevier-North Holland eds.), a chapter on "Forecasting in Macroeconomics" for the *Handbook of Research Methods and Applications in Empirical Macroeconomics*, and an article for the *Journal of Economic Literature*.

CREi
Centre de Recerca
en Economia Internacional

Academic unit of the
barcelona gse
graduate school of economics

Ramon Trias Fargas, 25-27 - 08005 Barcelona
Tel: 93 542 13 88 - Fax: 93 542 28 26
E-mail: crei@crei.cat
http://www.crei.cat

upf. Universitat Pompeu Fabra *Barcelona*

Generalitat de Catalunya