

Víctor López-Pérez

Macroeconomic Forecast Uncertainty in the Euro Area

Equilibrium. Quarterly Journal of Economics and Economic Policy 11/1, 9-41

2016

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.



López-Pérez, V. (2016). Macroeconomic Forecast Uncertainty in the Euro Area. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(1), pp. 9-41, DOI: <http://dx.doi.org/10.12775/EQUIL.2016.001>

Víctor López-Pérez *

Universidad Politécnica de Cartagena, Spain

Macroeconomic Forecast Uncertainty in the Euro Area

JEL Classification: D80; D84

Keywords: *macroeconomic uncertainty; Gini index; performance-based weights; Survey of Professional Forecasters; European Central Bank*

Abstract: *This paper estimates aggregate measures of macroeconomic uncertainty from individual density forecasts by professional forecasters. The methodology used in the paper improves on the existing literature along two dimensions. Firstly, it controls for changes to the composition of the panel of respondents to the survey. And secondly, it assigns more weight to the information submitted by forecasters with better forecasting performance. Using data from the European Central Bank's Survey of Professional Forecasters from 1999 Q1 to 2014 Q3, the paper finds that the effects of changes in the composition of the panel on aggregate uncertainty can be large in a statistical and economic sense. It also finds that the estimates of aggregate uncertainty that use performance-based weights differ significantly from the simple averages used in the literature and their dynamics are more consistent with the dynamics displayed by the estimates of uncertainty computed from financial indicators.*

© Copyright Institute of Economic Research

Date of submission: November 8, 2015; date of acceptance: January 19, 2016

* Contact: victor.lopez@upct.es, Universidad Politécnica de Cartagena, Department of Economics, c/ Real 3, Office 234, ZIP code: 30201, Cartagena, Spain

** I thank Benjamin Beckers, İsmail Hakki Genç, Maximilian Podstawski, Oreste Tristani, Kenneth Wallis, participants at the 1st Annual International Conference on Social Sciences (Istanbul, 2015), the 10th Warsaw International Economic Meeting (Warsaw, 2015) and the Lupcon Finance and Economics Conference (Frankfurt, 2015), and two anonymous referees for helpful comments on previous versions of this paper.

Introduction

Macroeconomic outcomes are the result of millions of decisions taken by economic agents worldwide, and the economic literature tries hard to understand the determinants of these decisions. One of those determinants is the degree of uncertainty in the economy, which is especially important for savings and investment decisions. If uncertainty is large, consumers are expected to save more for precautionary reasons (Caballero, 1990, pp. 113-136) and risk-averse investors may delay irreversible investment plans (Leahy & Whited, 1996, pp. 64-83).

The end of the Great Moderation era has put the term ‘uncertainty’ back to the front pages of newspapers and academic articles (Baker *et al.*, 2015, pp. 1-73). Uncertainty played a major role in the freeze of credit markets worldwide after the fall of Lehman Brothers in 2008. It was also the key factor in the propagation or contagion of the sovereign debt crisis from Greece to other European countries in 2011.

The evolution of macroeconomic uncertainty over time is thereby of interest for academics and researchers, who need estimates of uncertainty to investigate the links between uncertainty and economic outcomes. It is also of interest for policy-makers, who need to closely monitor the available estimates of uncertainty, anticipate the effects on the economy of changes in uncertainty and take the appropriate policy actions to achieve their policy objectives (Bloom, 2009, pp. 623-685).

The first place to look at for measures of risk and proxies for uncertainty is probably the financial market. The Chicago Board Options Exchange Volatility Index (VIX) is a well-known example (Basu & Bundick, 2014, pp. 1-56). However, the non-conventional policies by many governments and central banks in developed economies during the recent past may have involuntarily contributed to distort the signals extracted from financial markets in general and proxies for uncertainty in particular (Bekaert *et al.*, 2013, pp. 771-788).

In this context, data from surveys may provide a more accurate picture of macroeconomic uncertainty than financial indicators. The density forecasts of macroeconomic variables by professional forecasters are particularly valuable for the estimation of the degree of macroeconomic uncertainty perceived by survey participants. They combine the expertise of highly-skilled professionals with the heterogeneity of views that naturally comes from survey methods. Consequently, the measures of uncertainty from surveys like the Federal Reserve Bank of Philadelphia’s Survey of Professional Forecasters, the European Central Bank’s Survey of Professional Fore-

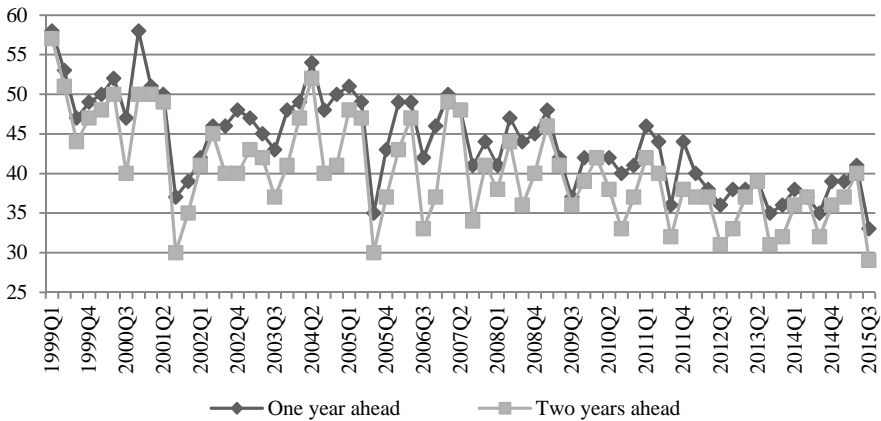
casters (ECB's SPF) and the Bank of England's Survey of External Forecasters have gained prominence in economic and policy discussions.

The existing literature has explored different measures of uncertainty constructed from survey data. The available measures of uncertainty are the standard deviation of point forecasts, typically known as "disagreement" (Neamtiu *et al.*, 2014, pp. 1071-1099), the variance of the average density forecast (ECB, 2014, pp. 55-67), the average standard deviation of the individual density forecasts (Giordani & Söderlind, 2003, pp. 1037-1059), the root mean subjective variance or RMSV (Batchelor & Dua, 1996, pp. 333-341), the implied RMSV (Boero *et al.*, 2008, pp. 1107-1127), the average or median inter-quartile range of the individual density forecasts (Engelberg *et al.*, 2011, pp. 1059-1078; Abel *et al.*, 2015), and the average entropy of the individual density forecasts (Rich & Tracy, 2010, pp. 200-207; Wallis, 2006, pp. 1-2).

All these measures exhibit two problems. The first is the *panel-composition problem*: the existing measures of aggregate uncertainty do not take into account that the panel of professional forecasters changes from one survey to another. Therefore, the evolution of these measures of uncertainty over time may be meaningless, because it compounds true changes in uncertainty with artificial changes due to the variations in the panel of respondents to the survey (Engelberg *et al.*, 2011, pp. 1059-1078).

Figure 1 shows the number of professional forecasters that submitted density forecasts of GDP growth one and two years ahead in each ECB's SPF round. There are large variations in the pool of respondents from one survey round to the next. For instance, the ECB received 49 density forecasts of GDP growth two years ahead in 2005 Q2. In the following round, 2005 Q3, it received only 35. Let us assume that there were no true changes in uncertainty between 2005 Q2 and 2005 Q3. Let us also assume that the forecasters that replied in 2005 Q2 but not in 2005 Q3 perceived less (more) uncertainty on average than the forecasters that participated in 2005 Q2 and 2005 Q3. Then, the estimates of aggregate uncertainty in 2005 Q3 would increase (decrease) not because uncertainty has truly changed, but because the ECB lost almost one third of its panel in 2005 Q3.

Figure 1. Number of ECB's SPF participants that submitted density forecasts of GDP growth one and two years ahead in each survey round



Source: ECB's SPF.

Abel *et al.* (2015) partially addressed the panel-composition issue by dropping from their sample the panellists with less than a minimum number of responses. They considered five different scenarios, with the minimum number of responses by panellist set at 5, 10, 15, 20 or 25. While this procedure may reduce the impact of panel changes, their measures of uncertainty still mix data from panellists with different participation patterns and, as a result, are still contaminated by variations in the panel of respondents to the survey.

The second problem the existing measures of aggregate uncertainty used in the SPF literature suffer from is the same-weight problem: they all assign the same weight to the uncertainty perceived by each survey participant, without taking forecasting performance into consideration. There are reasons to believe that the simple average of the individual measures of uncertainty may be different than a weighted average based on forecasting performance. Kenny *et al.*, (2015, pp. 1203-1231) found a positive relationship at the individual level between forecasting performance and the variance of the density forecasts in the ECB's SPF: the best-performing respondents submitted density forecasts with more probability in the tails (i.e. they perceived more uncertainty). If the highest weights are given to the best-performing forecasters when aggregating their individual estimates of uncertainty, the resulting estimates of aggregate uncertainty may differ from the unweighted estimates used in the SPF literature.

I am not aware of any study that used performance-based weights to construct estimates of aggregate uncertainty from the degree of uncertainty perceived at the individual level. The literature on forecast combination has explored the use of performance-based weights to obtain a weighted average of individual point forecasts (Stock & Watson, 2004, pp. 405-430; Capistrán & Timmermann, 2009, pp. 428-440; Smith & Wallis, 2009, pp. 331-355; Genre *et al.*, 2013, pp. 108-121). Others have used weights to obtain an aggregate density forecast as a combination of individual density forecasts (Hall & Mitchell, 2005, pp. 995-1033; Hall & Mitchell, 2007, pp. 1-13; Geweke & Amisano, 2010, pp. 130-141; Jore *et al.*, 2010, pp. 621-634; Kascha & Ravazzolo, 2010, pp. 231-250; Billio *et al.*, 2013, pp. 213-232; Conflitti *et al.*, 2015, pp. 1096-1103). However, performance-based weights have never been used for the aggregation of the estimates of uncertainty perceived by each survey participant into an estimate of aggregate uncertainty.

This paper solves these two problems and is organized as follows. The next section presents the methodology used in the paper and summarizes the main results. The following section introduces the ECB's SPF and describes the data used in the analysis. Then, the paper discusses how SPF-based measures of uncertainty that do not control for changes to the panel of respondents are affected by its unbalanced nature, and proposes a practical solution to this problem. Next, the paper presents the estimates of aggregate uncertainty when the individual measures of uncertainty are weighted according to each participant's forecasting performance. Finally, the paper concludes, discusses the implications of its findings and proposes directions for future research.

Method of the Research

This paper solves the panel-composition problem by estimating an aggregate measure of uncertainty from the data submitted by forecasters that replied to two consecutive survey rounds. Using the data from the ECB's SPF from 1999 Q1 to 2014 Q3, the paper finds that the effects of changes to the composition of the panel on aggregate uncertainty can be large in a statistic and economic sense. Moreover, changes to the composition of the panel may alter the direction of change of aggregate uncertainty measures. In this regard, the current standard of aggregating the results from all the participants in each survey round, independently of their participation in previous rounds, may produce very misleading results.

The paper then solves the same-weight problem by using performance-based weights to obtain weighted averages of individual measures of uncertainty. Forecasting performance is assessed with a strictly proper scoring rule: the logarithmic score of the density forecasts (Gneiting & Raftery, 2007, pp. 359-378). The forecasters with better scores are assigned higher weights. The paper finds that, in the ECB's SPF sample, the weighted estimates of aggregate uncertainty differ significantly from the simple averages used in the literature. The differences are statistically significant and economically relevant. In particular, weighted estimates indicate a much larger increase in uncertainty than the simple averages since the start of the financial crisis. Moreover, while the unweighted estimates of aggregate uncertainty have stayed rather flat in the euro area since 2010, the weighted estimates display significant variation. The latter are much more consistent with the shocks that have hit the euro area in the last five years, like the sovereign debt crisis and the recession in 2012. They are also much more in line with the volatility displayed by the uncertainty indicators from financial markets, like the VSTOXX index.

The Data

The ECB's SPF

The ECB's SPF surveys expectations by professional forecasters located in the European Union each quarter since 1999 Q1. Most forecasters work for financial institutions while others belong to universities, government agencies, labor unions and business organizations. 103 forecasters have participated at least once in the survey, although average participation is around 60 forecasters per round. The panel is unbalanced: many forecasters do not reply sometimes, while others have left the panel and have been replaced with new panellists.

The ECB's SPF surveys density forecasts of the year-on-year inflation rate, the year-on-year GDP growth rate, and the unemployment rate, all for the euro area. The forecasts used in this paper are fixed-horizon expectations, in particular, expectations one year ahead.¹ Fixed-event forecasts (e.g. expectations for the "current calendar year" or the "next calendar year") are not used because uncertainty is expected to shrink mechanically

¹ The results for expectations two years ahead are qualitatively similar to the results presented in the paper. These results are available from the autor upon request.

if the forecast horizon shortens from one survey round to the next.²

Treatment of open-ended bins

In order to produce a density forecast, forecasters are asked to distribute probabilities among a set of predefined bins for each variable.³ The lowest and the highest bins are open ended (e.g. “less than 0%” or “more than 3.9%”). For the measurement of uncertainty, an assumption needs to be made on the probability placed in open-ended bins. These bins are much less informative than the rest because of their infinite width. Previous studies have typically assumed that open-ended bins have the same or double width than closed bins (e.g. Batchelor & Dua, 1996, pp. 333-341). This assumption may lead to underestimate uncertainty, especially if open-ended bins contain relatively large probabilities. For instance, forecaster 52 in round 2009 Q1 assigned all the probability to GDP growth being “less than -1.0%” in one year. Given that the width of the closed bins in the ECB’s SPF is 0.5%, it seems hard to believe that he/she would have assigned all the probability either to the [-1.5%,-1.1%] bin or to the [-2.0%,-1.1%] bin when his/her point forecast was -2.9%.

To avoid drawing wrong inference from uninformative data, any density with at least one open-ended bin that cannot reasonably have the same width than the closed bins is removed from the sample. Specifically, any density with 50% more probability in an open-ended bin than in any other bin with non-zero probability is excluded from the sample used in this paper: 1999 Q1–2014 Q3.⁴ An exception is made for the densities with less than 1% probability in open-ended bins. Otherwise, all computer-generated densities with support from $-\infty$ to ∞ (e.g. a normal density function) would be excluded.

For the remaining density forecasts, open-ended bins are treated as having the same width than closed bins. Thus, the next sections do not make any distinction between closed and open-ended bins: all bins are assumed to be closed and assumed to have the same width.

² See <http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html> for a full description of the survey.

³ Details on the bins available to the SPF forecasters and on the forecast horizons surveyed in each SPF round can be obtained from the document “ECB Survey of Professional Forecasters (SPF): description of SPF dataset”, available here: http://www.ecb.europa.eu/stats/prices/indic/forecast/shared/files/dataset_documentation_csv.pdf?78b0b9ba730b2241d43fec92dacd2944d.

⁴ For completeness, densities with 100% probability in open-ended bins are also excluded.

The choice of the measure of individual uncertainty

The measure used to estimate the degree of uncertainty perceived by each forecaster is new: the Gini index of the individual density forecast. Borrowed from the literature on income and wealth inequality, the Gini index (Gini, 1955, pp. 211-382) is based on the Lorenz curve (Lorenz, 1905, 209-219). This curve is typically used to represent the percentage of total wealth in the hands of the poorest $x\%$ of the population. The Lorenz curve may also be applied to the analysis of uncertainty by representing the cumulative probability allocated to the x least likely bins of a density forecast.

If a forecaster faced no uncertainty, her density forecast would have probability 1 in one bin and 0 in the rest. In this case, the Lorenz curve would be 0 from the first bin to the one before the last and then it would jump to 1 in the last bin. If a forecaster faced maximum uncertainty, her density forecast would have the same probability allocated to every bin. Then, the Lorenz curve would increase uniformly from the first bin to the last.

In this context, the individual Gini index of uncertainty of a given density forecast is equal to the area between the Lorenz curve under maximum uncertainty and the Lorenz curve of the density forecast, divided by the area below the former (Gini, 1955, pp. 211-382). As the original Gini index would decline with uncertainty, it is multiplied by -1 in this paper to make it an index that increases with uncertainty.

The Gini index has two advantages over the most frequently used measure of uncertainty from density forecasts, the standard deviation of the individual density forecast (Batchelor & Dua, 1996, pp. 333-341; Giordani & Söderlind, 2003, pp. 1037-1059; Boero *et al.*, 2008, 1107-1127). Firstly, the Gini index takes its maximum value when the density forecast is uniform, i.e. when the forecaster faces maximum uncertainty and all the bins are perceived as equally likely. The standard deviation of a density forecast reaches its maximum when the forecaster puts 0.5 probability in the lowest bin and the other 0.5 in the highest. Obviously, the formulation of the latter density forecast requires a lot of information, e.g. that the probability of an outcome located in the intermediate bins is zero. This amount of information is completely at odds with the notion of maximum uncertainty.

The second advantage of the Gini index over the standard deviation of a density forecast is that the computation of the standard deviation requires an assumption on how the probability is distributed within each bin.⁵ Some

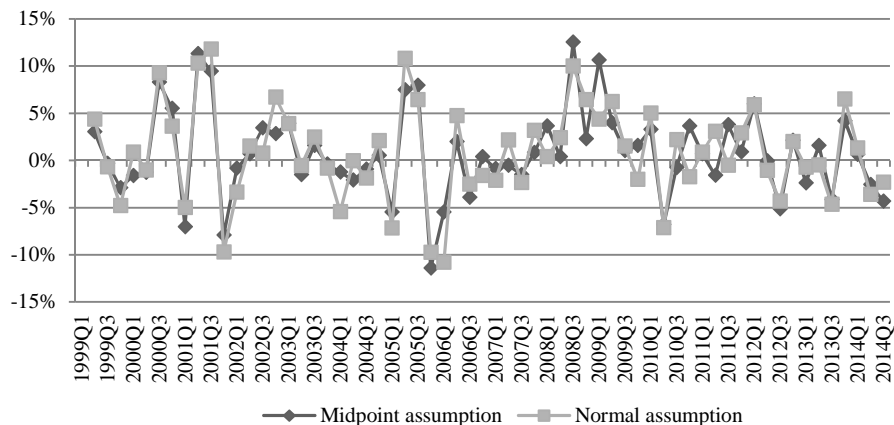
⁵ Another popular measure of uncertainty in the SPF literature, the inter-quartile range of the density forecast, suffers from the same problem.

studies assume that all the probability allocated to a bin is in the middle point of the bin (Rich & Tracy, 2010, pp. 200-207) while others fit a continuous distribution to the histogram. The normal distribution (Boero *et al.*, 2015) and the generalized beta distribution (Engelberg *et al.*, 2011, pp. 1059-1078; Abel *et al.*, 2015) are the most commonly used while the piecewise linear distribution (Conflitti, 2011, pp. 69-103) and the skew-normal distribution (García & Manzanares, 2007, pp. 1-47) have also been proposed.

How robust is the standard deviation to these assumptions? I computed the square root of the average variance (the so-called RMSV, Batchelor & Dua, 1996, pp. 333-341) of the individual density forecasts of inflation two years ahead under two different assumptions.⁶ Under the first assumption, all the probabilities are assumed to be located in the middle point of each bin. Under the second assumption, a normal density is fitted to the each individual density forecast. Figure 2 shows the RMSVs under these alternative assumptions, in percentage changes from the previous quarter. The two series are clearly correlated but they do not coincide. For example, in 2009 Q1 the RMSV increased by 10.6% under the middle-point assumption but only by 4.4% under the normal assumption. Interestingly, even the direction of change in the RMSV statistic may depend on the assumption made. For instance, in 2010 Q4 the RMSV increased by 3.6% under the middle-point assumption but declined by 1.7% under the normal assumption. This lack of robustness is undesirable. I would use these standard-deviation-based measures if all the alternative measures needed such assumptions, but there are some alternatives that do not need them.

⁶ Although the rest of the paper focuses on one-year-ahead forecasts, I chose to show the RMSVs of inflation forecasts two years ahead because they clearly portray how important this assumption becomes.

Figure 2. RMSV of individual density forecasts of inflation two years ahead under different assumptions regarding how the probabilities are distributed within each bin (percentage change from the previous quarter)

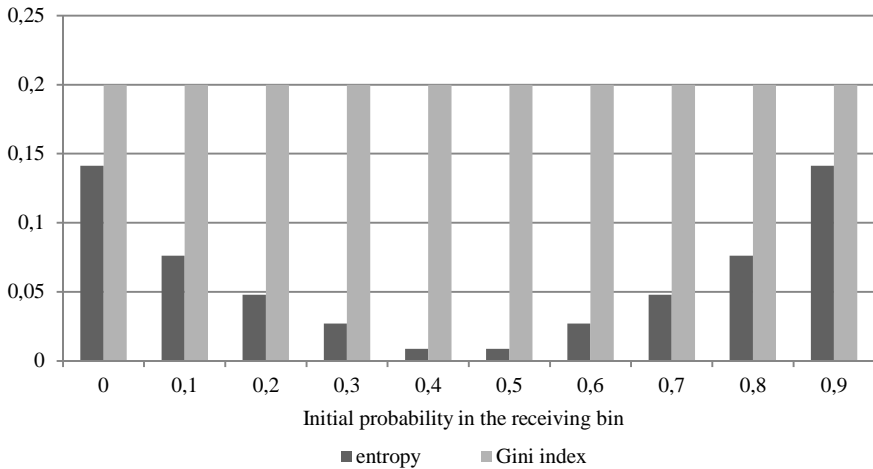


Source: own calculations based on ECB's SPF data.

The Gini index of the density forecast does not require any assumption about the distribution of the probability inside each bin, making it more attractive than the standard deviation of the density forecast as a measure of individual uncertainty. An alternative measure that does not require this assumption either is the entropy of the individual density forecast (Rich & Tracy, 2010, pp. 200-207; Wallis, 2006, pp. 1-2). Why not use the entropy of a density forecast to measure uncertainty? As the Gini index, the entropy takes its maximum value when the density forecast is uniform and it does not need any assumption regarding the distribution of the probabilities within each bin. However, the Gini index has an advantage over the entropy: the non-linear nature of the entropy implies that the effect on the entropy from a certain change in the probabilities of a density forecast depends on the initial values of these probabilities. In the context of a simple example with only two bins, Figure 3 shows that moving 0.1 probability from one bin to the other leads to larger absolute changes in the entropy when the initial probabilities allocated to the two bins are very different, i.e. when the level of entropy is smaller. The Gini index does not suffer from this drawback.⁷

⁷ As shown on Figure 3, this is always true with two bins. With three or more bins, the Gini index retains this property as long as the change in the probabilities does not alter the ordering of the bins in the Lorenz curve.

Figure 3. Illustration of the absolute changes in the entropy and the Gini index when 0.1 probability is moved across bins (example with two bins)



Source: own calculations.

This property of the entropy is relevant in the context of the ECB’s SPF because some forecasters assign zero probability to too many intervals. This behavior has been labeled “overconfidence”, and it worsens the forecasting performance (Kenny *et al.*, 2015, pp. 1203-1231).⁸ The entropy of the density forecasts submitted by overconfident forecasters is smaller than the entropy of the density forecasts submitted by more “prudent”, better-performing forecasters. As changes in the entropy are larger when the initial level of entropy is smaller, changes in the average entropy would be relatively more affected by changes in the behavior of the overconfident forecasters. Thus, the entropy implicitly puts more weight on the worst forecasters.

The Gini index is immune to this problem as well. For these reasons, I choose the Gini index of the individual density forecast as the estimate of the level of uncertainty perceived by each SPF forecaster.

Realisations of the forecasted variables

To weight the individual perceptions of uncertainty by each forecaster based on the forecaster’s performance, I need to compute scores by fore-

⁸ Overconfidence is defined here as “excessive precision in one’s beliefs” (Moore & Healy, 2008, pp. 502-517).

caster. To this end, I compare the density forecasts with the realizations of the forecasted variables. These realizations are retrieved from the ECB's Euro Area Real-Time Database.⁹ This database collects vintages of many macroeconomic variables as they appeared in each issue of the ECB Monthly Bulletin (Giannone *et al.*, 2010, pp. 1-119).

Real-time data is used in this paper because ECB's SPF participants tried to forecast inflation, GDP growth and unemployment as defined by the statistical methodology existing at the time of the production of the forecast. If the latest vintage of data were used instead, differences between forecasts and realizations would not only arise because of forecast errors. They may also be caused by subsequent methodological changes to the calculation of the forecasted variable that led to backward revisions in the historical time series of the variable.

At the time of retrieving the data from the Real-Time Database (July 2015), inflation data was available until September 2014, real GDP growth data until 2014 Q3, and unemployment data until August 2014. Due to the length of the forecast horizons in the ECB's SPF, scores may be computed up to 2013 Q4 for density forecasts of inflation one year ahead, 2014 Q1 for density forecasts of GDP growth one year ahead, and 2013 Q4 for density forecasts of unemployment one year ahead.

Solving the Panel-composition Problem

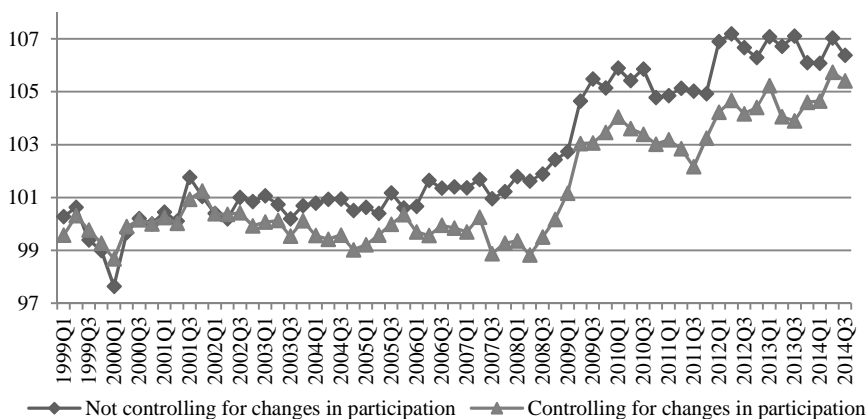
The SPF literature combines the measures of individual uncertainty obtained from the individual density forecasts into a measure of aggregate uncertainty. This aggregation is done by taking the mean or the median of the measures of individual uncertainty for all the forecasters that participated in a given survey round. By proceeding in this way, the existing estimates of aggregate uncertainty do not take into account that the SPF panel of professional forecasters changes from one survey round to the next. Engelberg *et al.* (2011, pp. 1059-1078) argued the evolution of these measures of uncertainty over time is mixing true changes in uncertainty with artificial changes due to the variations in the panel of respondents to the survey. In their analysis of the US SPF, they recommended going beyond aggregate figures and to examine the changes to the individual responses of the survey. This is the road this section takes, trying to ascertain the effects on the aggregate measures of uncertainty caused by the entry and exit of forecasters from the ECB's SPF panel.

⁹ <http://sdw.ecb.europa.eu/browse.do?node=4843526>.

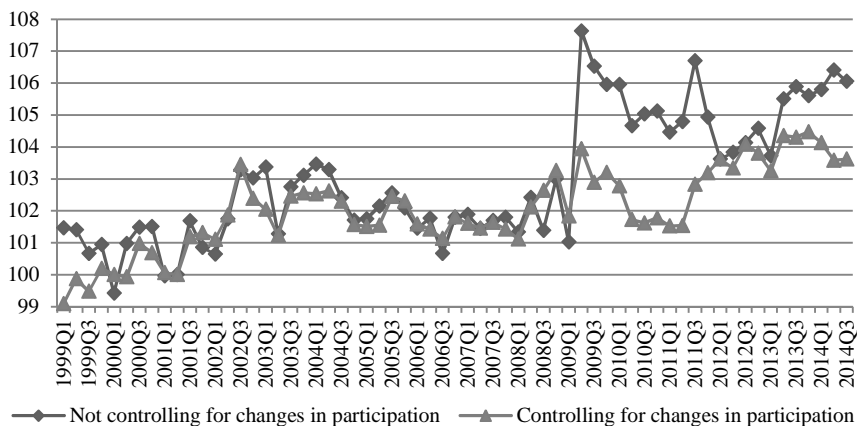
Figure 4 shows the average Gini indices of the density forecasts submitted by all the respondents to the ECB’s SPF (lines with diamonds). Each index has been normalized to 100 in a base quarter.¹⁰ These series are very similar to the ones used in the SPF literature and do not control for changes in the composition of the panel of respondents to the survey.

Figure 4. Estimates of aggregate uncertainty for the ECB’s SPF

a) From density forecasts of inflation one year ahead (2000 Q1 = 100)

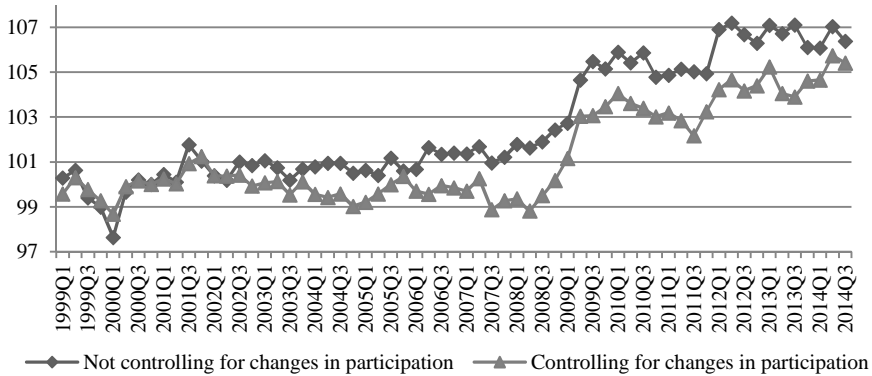


b) From density forecasts of GDP growth one year ahead (2001 Q2 = 100)



¹⁰ The base quarter is not the same for all panels in Figure 4. The reason will become clear in the next section.

c) From density forecasts of unemployment one year ahead (2000 Q4 = 100)



Source: own calculations based on ECB's SPF data.

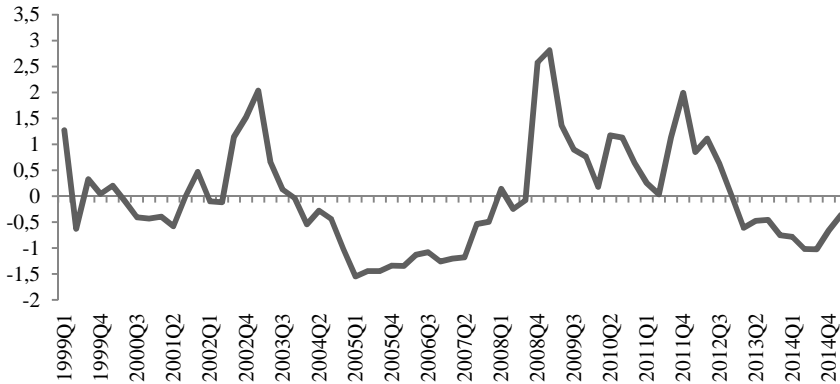
To control for entry to and exit from the pool of respondents, I select for each survey the subset of respondents that replied both to that survey and to the previous survey. Then, I compute the average Gini indices of the density forecasts submitted by this subset of forecasters in both survey rounds. The percentage change between these two indices is an estimate of how uncertainty has changed between the two survey rounds. This estimate is not affected by changes in participation by construction.

These percentage changes in uncertainty from one survey round to another make it possible to construct an index of aggregate uncertainty that is not affected by changes to the panel of respondents to the survey. These indices are the lines with triangles in Figure 4. A comparison between the lines with diamonds and triangles lines clearly shows that controlling for changes in the panel of respondents to the ECB's SPF is important for the estimation of macroeconomic uncertainty in the euro area. Look, for example, at panels a and c, where density forecasts of inflation or unemployment are used to estimate uncertainty. Aggregate uncertainty clearly falls from 2002 to 2007, when I control for entry and exit, while uncertainty does not decline by much or at all when I use the data from all the respondents to the survey. The period from 2002 to 2007 was one of economic bonanza, characterised by declining uncertainty and volatility: Figure 5 shows the standardised 12-month VSTOXX index of financial market volatility. This index is the euro-area version of the VIX index, which has been used to proxy for macroeconomic uncertainty.¹¹ The VSTOXX index was much lower in

¹¹ The VSTOXX indices are based on Eurostoxx 50 real-time options prices and are designed to reflect the market expectations of short-term and long-term volatility by measuring the square root of the implied variance across all options of a given time to expiration. The

2007 than in 2002, more in line with the estimates of macroeconomic uncertainty from the ECB’s SPF that control for entry and exit.

Figure 5. Standardised 12-month VSTOXX index of financial-market volatility from European financial markets



Source: own calculations based on stoxx.com data.

Table 1. Correlation between the changes in the aggregate Gini index of uncertainty using the complete pool of respondents and using the subset of respondents that submitted their density forecasts during two consecutive rounds

Variable	Correlation coefficient (1999:2-2014:3)
Inflation	0.67
GDP growth	0.71
Unemployment	0.64

Source: own calculations based on ECB’s SPF data.

Table 1 displays the correlation coefficient between changes in the aggregate Gini indices of uncertainty using the complete pool of respondents, and those using the subset of respondents that submitted density forecasts during two consecutive rounds. The correlations are not close to 1, ranging from 0.64 to 0.71. Therefore, the use of a measure of uncertainty that aggregates the results from all the respondents may lead to misleading results

data is obtained from http://www.stoxx.com/download/historic_al_values/h_vstoxx.txt. The quarterly data shown on Figure 5 are average daily data over the quarter.

because some of its variation is caused not by true changes in uncertainty but by changes in the composition of the panel of respondents.

Solving the Same-weight Problem

The measures of aggregate uncertainty used in the SPF literature are unweighted averages that assign the same weight to the uncertainty perceived by each survey participant, without taking forecasting performance into consideration. They do not give more weight to the information submitted by a forecaster who always performs better than the average, or less weight to the responses by a forecaster that always underperforms. As a result, the unweighted estimates of aggregate uncertainty used in the literature may be biased because they may implicitly give a weight larger than optimal to underperforming forecasters.

Weighted averages of individual estimates of uncertainty

This section presents estimates of aggregate uncertainty computed with data from the ECB's SPF. These estimates are weighted averages of individual measures of uncertainty, instead of the simple unweighted averages used in the literature. The weight assigned to each forecaster is based on his/her forecasting performance, which is assessed by the logarithmic scoring rule. This rule is one of the four strictly proper scoring methods described by Gneiting & Raftery (2007, pp. 359-378).

The logarithmic score is one of, if not the most popular, scoring rule (see Hall & Mitchell, 2005, pp. 995-1033; Kascha & Ravazzolo, 2010, pp. 231-250, for examples of uses of the logarithmic score with survey data). The logarithmic score by forecaster i in period t is:

$$S_{it} = \log p_{rit} \quad (1)$$

where p_{rit} is the probability assigned by the forecaster to the bin that includes the realization of the forecasted variable. The logarithmic score takes a value of zero if the forecaster assigned all the probability to the bin where the realization fell and takes a value of minus infinity if the forecaster placed zero probability in that bin. Each forecaster has a different score for each forecasted variable and forecast horizon, i.e. the score for inflation forecasts one year ahead by forecaster 1 is likely to differ from his/her score for inflation forecasts two years ahead.

Individual weights will be assigned on the basis of each forecaster's average performance over a period of time. Average performance by forecaster i at time t is computed as the simple average of his/her individual scores from period $t-W$ to period t ,

$$\bar{S}_{it} = \frac{1}{1+W} \sum_{w=0}^W S_{it-w}, \quad W \geq 0 \quad (2)$$

For instance, if $W=1$, the individual average score is computed over the current and the previous survey rounds. Obviously, the individual average score cannot be computed for the first W survey rounds. If a forecaster has not participated in all the $1+W$ survey rounds, the average is computed over the rounds when he/she participated. The weight assigned to forecaster i in survey round t is assumed to be:¹²

$$W_{it} = \frac{e^{\bar{S}_{it-h}}}{\sum_{j=1}^J e^{\bar{S}_{jt-h}}} \quad (3)$$

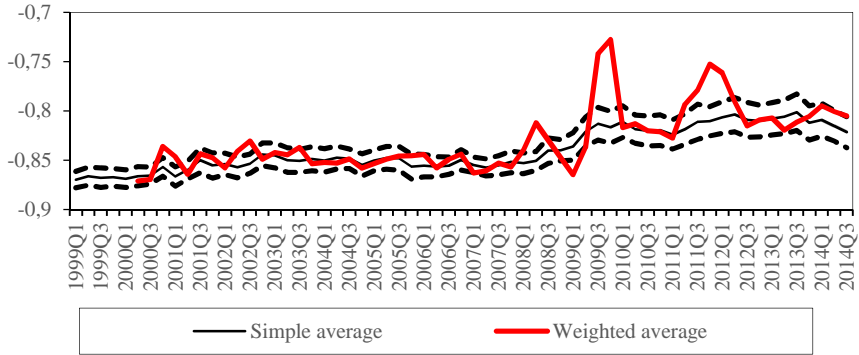
where J is the number of forecasters that participated in survey round t . Equation (3) implies that the sum of the weights of the participating forecasters equals one. h is the length of the forecast horizon (in quarters) of the variable of interest ($h=4$ for one-year-ahead forecasts). This guarantees that the weights can be computed in real time because the scores are available after four quarters for one-year-ahead forecasts.

Intuitively, the best forecasters, i.e. those with the highest average logarithmic scores, will receive the highest weights. Those weights may then be used to obtain an estimate of aggregate uncertainty equal to the weighted average of the individual Gini indices of the density forecasts. This is the point at which this paper departs from the existing SPF literature, which uses simple averages. Figure 6 shows the resulting estimates of aggregate uncertainty for the expected inflation rate, the expected rate of growth of real GDP and the expected unemployment rate surveyed in the ECB's SPF.

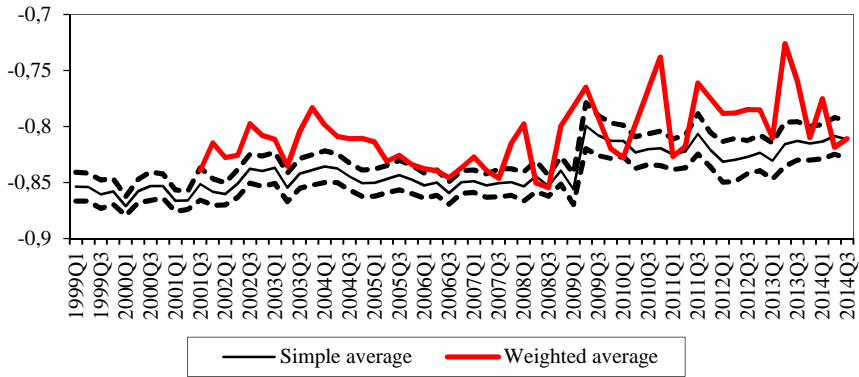
¹² For the estimation of aggregate uncertainty under optimal weights, see the next subsection.

Figure 6. Simple and weighted averages of individual measures of uncertainty

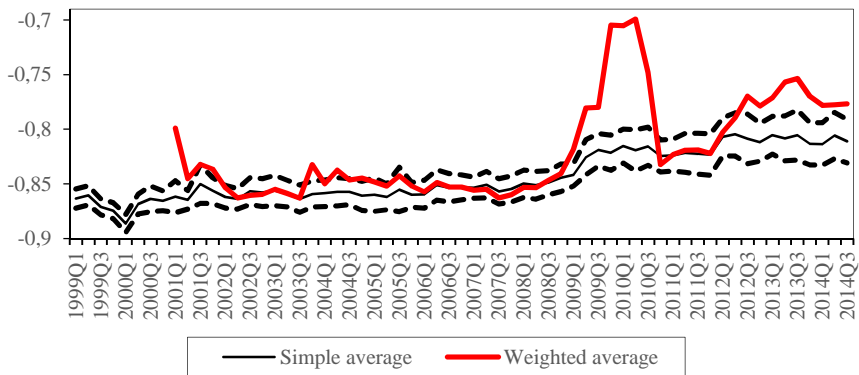
a) From density forecasts of inflation one year ahead



b) From density forecasts of GDP growth one year ahead



c) From density forecasts of unemployment one year ahead



Source: own calculations based on ECB's SPF data.

Each chart shows a solid black line, which is the time series of the simple average of the individual Gini indices of uncertainty. The two dashed lines are the bounds of the 95% confidence interval around the simple average. The red line is the weighted average of the individual Gini indices of uncertainty. Note that I use here the data submitted by all forecasters, and thereby I am not controlling for changes in the composition of the panel of participants. I will control for entry and exit below.

For the estimates of uncertainty based on density forecasts of inflation, the weighted and unweighted estimates are similar until the start of the financial crisis. Then, the weighted estimates suggest a much larger increase in uncertainty than the unweighted estimates. After the initial increase, the unweighted estimates stay rather flat, while the weighted estimates indicate a sharp decrease in uncertainty, followed by another, more moderate increase around 2011–2012 when fears of a potential break-up of the euro area mounted. After the president of the ECB announced in the summer of 2012 that the institution would do “whatever it takes” to preserve the integrity of the euro area, the weighted estimates of uncertainty declined significantly. None of these movements can be observed in the unweighted estimates.

The estimates of uncertainty based on density forecasts of unemployment show similar dynamics. The unweighted estimates stay rather flat after the initial increase in 2008–2009. The weighted estimates show more interesting dynamics, with uncertainty declining sharply in 2010 when GDP in the euro area started to grow around 2% a year. Uncertainty increased again in 2012, when the euro area economy went into a recession. Finally, the weighted estimates of uncertainty fell again in 2014, when the GDP growth climbed above zero again. Weighted estimates of uncertainty based on density forecasts of GDP are more volatile.

Interestingly, the weighted estimates of uncertainty track much better than the unweighted ones the dynamics of uncertainty extracted from financial market indicators. Figure 5 showed the standardised 12-month VSTOXX index of stock market volatility in the euro area. The M-shaped dynamics of this index between 2008 and 2013 are remarkably similar to those of the weighted estimates of uncertainty described above. The unweighted estimates are not able to replicate these fluctuations.

The main results described above are robust to the value of W . I have tried with values ranging from zero to eleven quarters. Very high values of W increase the chances of giving zero weight to most forecasters, while low values of W induce more volatility in the individual weights over time. The values used in Figure 6 try to balance these two effects: W is set equal to 1 for inflation forecasts, 4 for unemployment forecasts, and 6 for GDP-

growth forecasts. The results for different values of W are available from the author upon request.

Even with these relatively low values of W , there are a few occasions when the average score is minus infinity for all the participants in a survey round. These episodes are concentrated around the start of the financial crisis, which most participants did not foresee. The weighted estimates of uncertainty based on the logarithmic score cannot be computed in such occasions, and are replaced by linear interpolations between the previous and the next weighted estimates available. Weighted estimates based on different proper scoring rules, like the Brier score, the rank-probability score or the spherical score do not suffer from this drawback and the main results of the paper are robust to the scoring rule used. The results based on these other scoring rules are available from the author upon request.

An interesting feature of the weighted estimates of aggregate uncertainty is that their volatility is higher than the volatility of the unweighted estimates. This is because of two reasons. First, relative performance is changing over time, especially during the most turbulent periods. During those periods the best forecasters outperform the rest. Consequently, the weights assigned to the best forecasters tend to increase in more turbulent times. In those periods, the weighted estimate of aggregate uncertainty deviates more from the simple average and moves closer to the individual measures of uncertainty by the best forecasters, which tend to be higher.

This effect can be seen in Figure 7, which shows the weights assigned to the best forecaster, who may change over time. The survey rounds when a weight is not shown are the rounds when the individual average logarithmic scores h quarters before were minus infinity for all the participants. In these cases, the denominator in equation (3) is zero.

The charts show that the weight assigned to the best forecaster increases in more turbulent periods, for example at the start of the financial crisis. These are the periods when the weighted estimate of aggregate uncertainty tends to deviate the most from the simple average (e.g. in 2009).

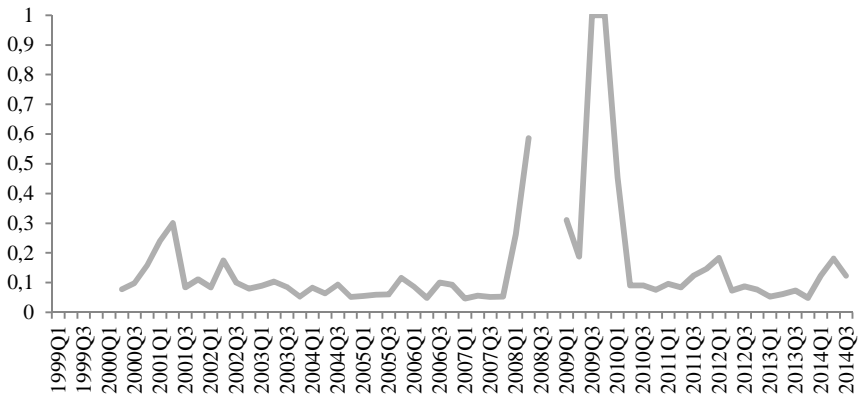
The second reason for the higher volatility of the weighted estimates compared to the unweighted estimates is that changes in participation from one round to the next cause changes in the weights. This is because the denominator in equation (3) varies, even if relative performances did not change much.

For instance, Figure 7 shows that the volatility of the weights is the highest for GDP growth forecasts (panel b), which translates into more volatile weighted estimates of uncertainty (panel b in Figure 6). On the contrary, the volatility of the weights is the lowest for inflation forecasts (panel a in Figure 7), which leads to more stable weighted estimates of

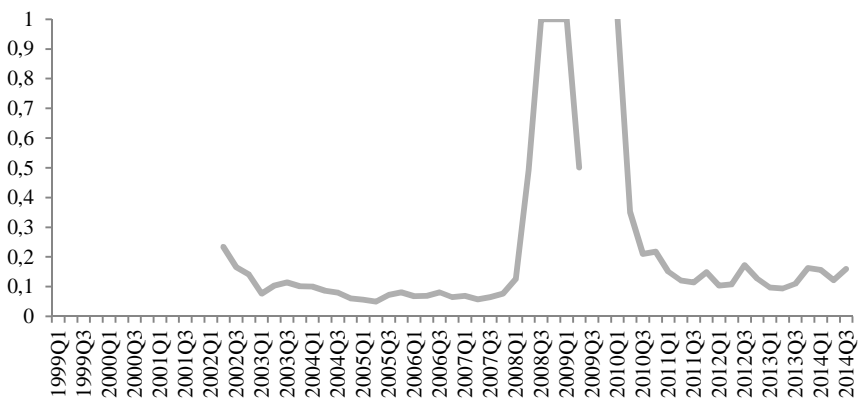
uncertainty (panel a in Figure 6). To eliminate the mechanical effects from changes in participation on the weights and then on estimates of uncertainty, I shall replicate the analysis using data from subsets of forecasters with the same participation patterns. As mentioned above, the results of this exercise will be presented below.

Figure 7. Weight assigned to the best forecaster

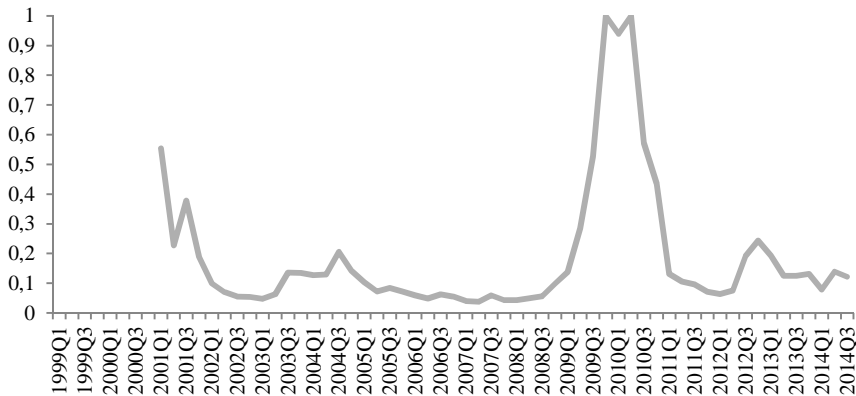
a) Using density forecasts of inflation one year ahead



b) Using density forecasts of GDP growth one year ahead



c) Using density forecasts of unemployment one year ahead



Source: own calculations based on ECB's SPF data.

Another interesting result of the comparison between weighted and unweighted estimates of aggregate uncertainty is that, when the former deviates from the latter, it is mostly to indicate higher uncertainty. Figure 6 shows how frequently the weighted estimate of aggregate uncertainty crosses the upper bound of the 95% confidence interval around the simple average. However, the weighted estimate very rarely crosses the lower bound of the confidence interval. In other words, simple averages of individual measures of uncertainty computed with data from the ECB's SPF may produce estimates of aggregate uncertainty that are frequently too low. The reason is that the same-weight rule assigns too much weight to "overconfident" forecasters. These forecasters' predictions are typically worse than the average forecast, reducing the signal-to-noise ratio of the estimations of aggregate uncertainty. When the weight assigned to these forecasters is lowered according to their forecasting performance, the estimates of aggregate uncertainty are frequently significantly higher, both statistically and economically.

Estimates of aggregate uncertainty with optimal weights

The previous subsection showed estimates of aggregate uncertainty with larger weights being given to the best forecasters. However, the functional form used for the calculation of the weights (equation (3)) was *ad-hoc*. This subsection checks if the main results presented above are robust to the use of optimal weights.

To compute the optimal weights, I follow Conflitti *et al.* (CDG, 2015, pp. 1096-1103) and minimize the Kullback-Leibler Information Criterion

(KLIC) between the true density function of the forecasted variable and the combined density forecast submitted by SPF forecasters,

$$KLIC = E \left[\ln q_v - \ln p_v^{spf} (W_1, W_2, \dots, W_I) \right] \quad (4)$$

where q_v is the true density function of variable v , and p_v^{spf} is the combined density forecast with time-invariant individual weights W_1, W_2, \dots, W_I . CDG assumed that the optimal weights are constant over time, which means that the optimal weights cannot be computed in real time, an important difference with respect to the previous subsection. The true density function is unobservable, but is independent of the weights. Therefore, the weights that minimize the KLIC are the weights that maximize the average logarithmic score of the combined density forecast.

The computation of the optimal weights requires a balanced panel. CDG replaced missing individual density forecasts with uniform densities. Unfortunately, this assumption makes non-respondents to perform better than many respondents, which results in disproportionately high weights for non-respondents. Instead, I removed from the sample the forecasters who had participated less than ten times since 1999 Q1. For the remaining forecasters, I assumed that the logarithmic score of the missing individual density forecasts in a given survey round is equal to the average logarithmic score of the density forecasts submitted by the remaining forecasters that participated in that survey round. This procedure is done separately for each forecasted variable (inflation, GDP growth and unemployment) because the optimal weights will be different for different variables.

Once the panel is balanced, I removed from the panel those forecasters whose logarithmic score is minus infinity in a period when at least another forecaster's score was better. This reduces the cross-sectional dimension of the optimization problem, and the maximization of the average logarithmic score would have assigned a weight equal to zero to these forecasters anyway. Interestingly, there are a few survey rounds when the logarithmic score is minus infinity for all the participating forecasters (e.g. in 2009 Q1 for GDP growth forecasts). This happens for instance when the realized value of the forecasted variable does not belong to any of the available closed bins. To avoid penalizing the forecasters who participated in a survey round with this technical deficiency of the questionnaire, I excluded these periods from the maximization problem.

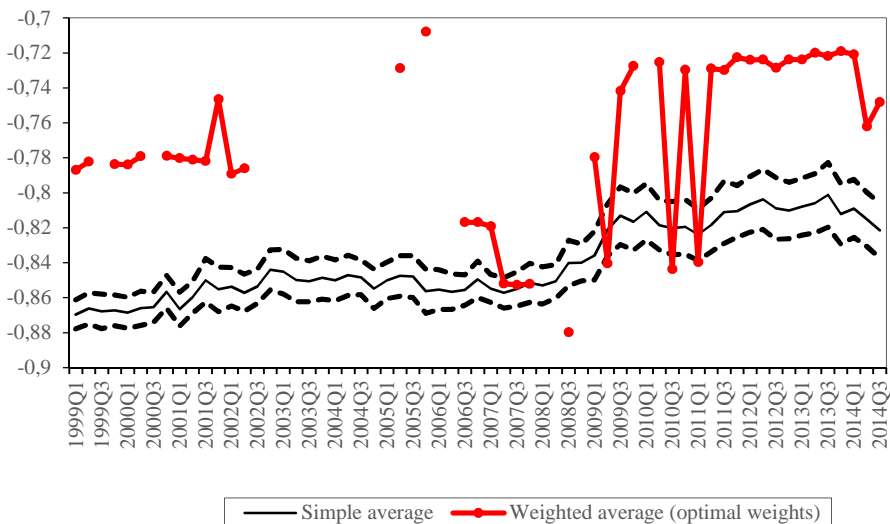
Two maximization routines are conducted. The first is a grid search, where the average logarithmic score of the combined density forecast is computed for every possible combination of non-negative weights that

sums one. The weights in the grid change in 0.01 steps. The second routine is the algorithm proposed by CDG. Both procedures yield the same results, and the number of non-zero elements in the vector of optimal weights turns out to be very small (two or three).

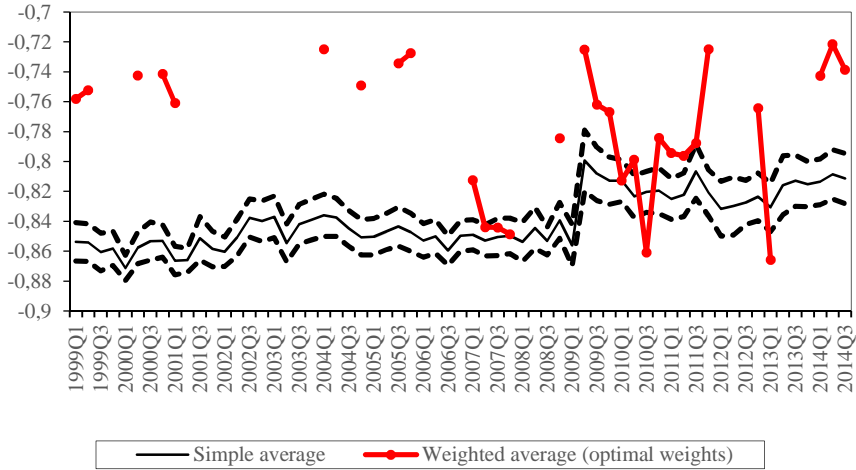
I now use the optimal weights to obtain the weighted averages of the individual estimates of uncertainty. Figure 8 shows the comparison between these weighted estimates and the simple unweighted estimates used in the literature. One thing to note is that there are some gaps in the weighted estimates. This happens when none of the participants with a positive optimal weight participated in a survey round (the panel was balanced for the optimization but is back to its unbalanced form for the estimation of uncertainty).

Figure 8. Weighted averages of individual measures of uncertainty using the optimal weights

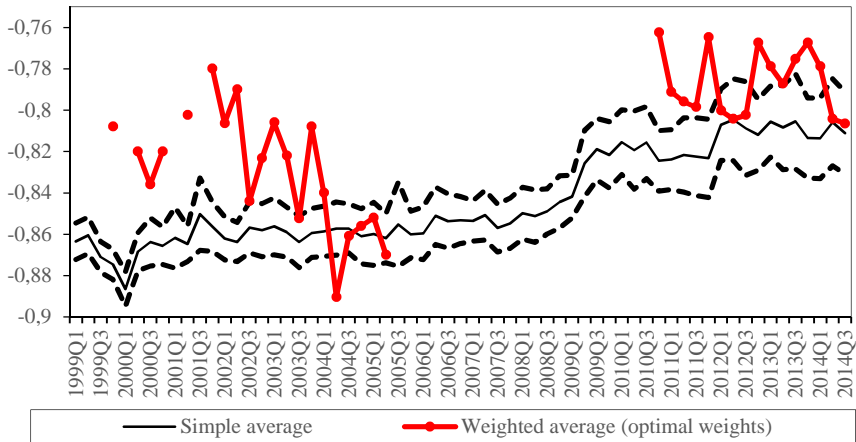
a) From density forecasts of inflation one year ahead



b) From density forecasts of GDP growth one year ahead



c) From density forecasts of unemployment one year ahead



Source: own calculations based on ECB’s SPF data.

Apart from this, the results obtained in the previous subsection are robust to the use of optimal weights. In particular:

- The weighted estimates of uncertainty increased by much more than the unweighted estimates at the start of the financial crisis. The estimated increase in uncertainty is several times larger using the weighted estimates.

- The unweighted estimates stayed relatively flat after their increase at the beginning of the financial crisis, while the weighted estimates of uncertainty display significant variation. For instance, panel b (estimates obtained from forecasts of GDP growth) shows that the weighted estimate of uncertainty declined significantly during 2009 and 2010 and then increased again in 2011. The unweighted estimates of uncertainty fail to produce such swings.
- The weighted estimates of uncertainty tend to suggest higher levels of uncertainty than the unweighted estimates. All the panels in Figure 8 show that deviations of the weighted average from the simple average are mostly to the upside. Importantly, these deviations are statistically significant and economically important. As described above, this result suggests that the best forecasters submit forecasts with more probability in the tails.

All in all, the results with the optimal weights go hand in hand with the main results presented in the previous subsection. But are these results still valid when the different participation patterns by SPF panellists are taken into account? The next subsection investigates this issue.

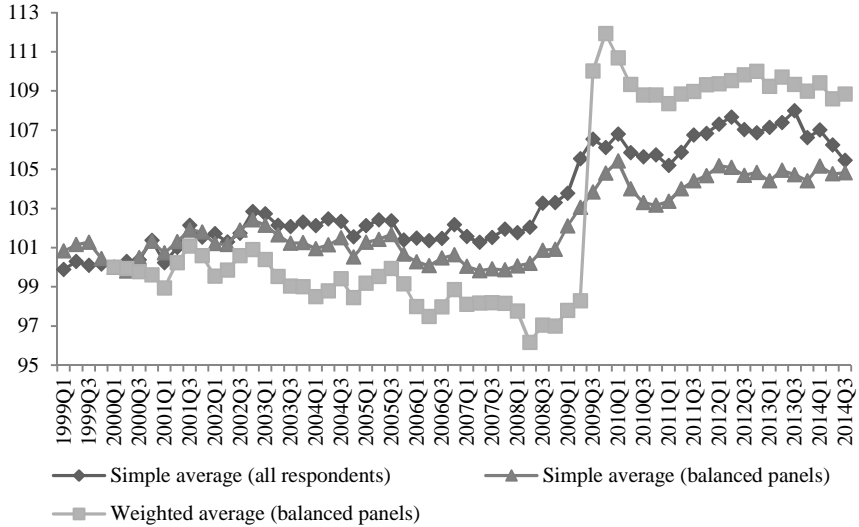
Estimates of aggregate uncertainty using subsets of forecasters with the same participation pattern

The analysis presented in the last two subsections has ignored the fact that the ECB's SPF panel dataset is unbalanced. In the previous section, I have showed that the estimated changes in aggregate uncertainty from one round to the next may be significantly contaminated by changes in participation. To control for entry and exit, I proposed to estimate changes in aggregate uncertainty between two rounds by focusing on the replies by the forecasters that have participated in both rounds. This is the approach followed in this subsection.

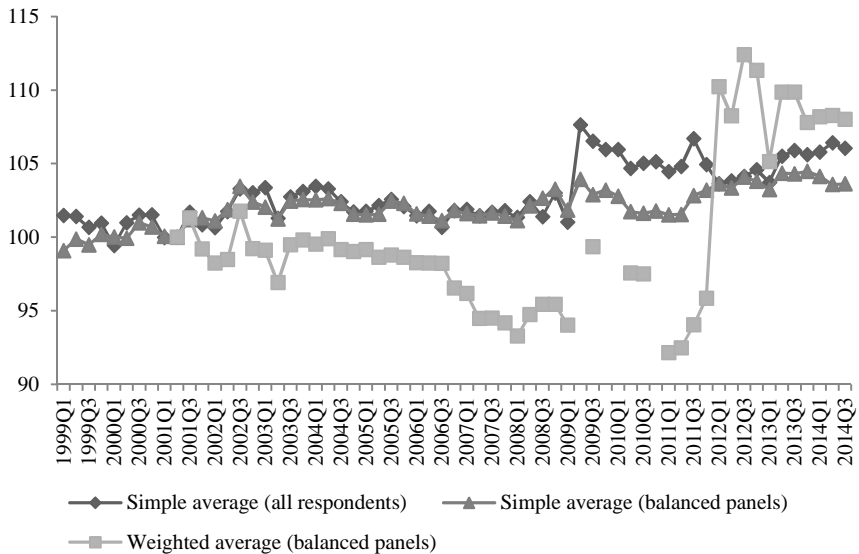
Figure 9 shows the resulting estimates of aggregate uncertainty when the same methodology described in the previous section is applied here. Each chart shows three lines. The line with diamonds is the unweighted average of the individual measures of uncertainty from all the respondents. It is the same index depicted in Figure 4. The lines with triangles and squares are two indices of aggregate uncertainty obtained from the replies by forecasters who participated in two consecutive rounds. The line with triangles is the index that assigns the same weight to each of these forecasters and is equal to the line shown on Figure 4. The line with squares depicts the index that assigns different weights according to forecasting performance.

Figure 9. Weighted averages of individual measures of uncertainty controlling for changes in participation.

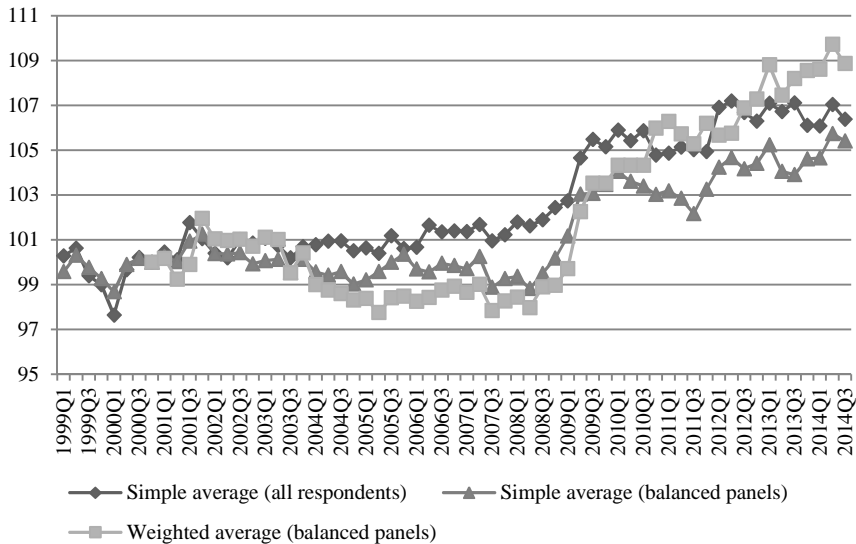
a) From density forecasts of inflation one year ahead (2000 Q1 = 100)



b) From density forecasts of GDP growth one year ahead (2001 Q2 = 100)



c) From density forecasts of unemployment one year ahead (2000 Q4 = 100)



Source: own calculations based on ECB's SPF data.

The weights used for the calculation of the weighted estimates of aggregate uncertainty (the lines with squares in Figure 9) are the real-time weights computed using equation (3). For the subset of forecasters who participated in rounds $t-1$ and t , the period- t weights are used to compute the weighted averages of their individual measures of uncertainty in periods $t-1$ and t . As stated above, the period- t weights are computed using the last $1+W$ quarters of available data. W varies across forecasted variables. Therefore, the weighted estimates of aggregate uncertainty start at different dates in each chart because survey data before 1999 Q1 is not available. The start date for the weighted estimates is defined as the base quarter for all the indices shown in each chart.¹³

The weights are re-scaled up to sum one, if necessary, because some forecasters who received a positive weight in round t may have not participated in round $t-1$. The gaps in the line with squares in Figure 9 (panel b) appear when zero weights are assigned to all the forecasters who participated in both rounds. In these cases, for the first value of the index after the gap, I used the data from the subset of forecasters who participated in the last round before the gap and in the first round after the gap.

¹³ This is the reason for the different base quarters across panels in Figure 4.

Four results are obtained from this analysis. First, the increase in aggregate uncertainty since the start of the financial crisis is estimated to be larger when higher weights are given to the best forecasters. For instance, the weighted estimate in panel a jumps by more than 14% between 2008 and 2009, while the unweighted indices did so by around 5% only. This finding confirms the results obtained in the previous subsections.

Second, the unweighted estimates tend to be rather flat after 2009, something not very realistic as explained above. The weighted estimates show significant variation after 2009. Again, this result confirms the findings presented in the previous subsections. For instance, the decline in the weighted estimates of aggregate uncertainty in the second half of 2009 and 2010 (see panels a and b) and their increase from the second half of 2011 (see panels b and c) match very well with the movements in the VSTOXX index shown on Figure 5.

Third, the weighted estimates of aggregate uncertainty computed from the forecasts of nominal variables (inflation) declined after 2009 and stabilized around relatively high levels. But the weighted estimates computed from the forecasts of real variables (GDP growth and unemployment) showed little signs of stabilization and continued to increase until 2014. Higher and higher uncertainty may help explain why the ECB took a series of non-conventional monetary-policy measures since 2010, which culminated in the announcement of the Public Sector Purchase Programme (or “Quantitative Easing”) in 2015. While these policy measures have calmed financial markets, as suggested by the sharp decline in the VSTOXX index since 2012, they do not seem to have had much success to reduce aggregate macroeconomic uncertainty, at least as perceived by professional forecasters. The developments in the unweighted estimates of aggregate uncertainty do not give any reason to take progressively more aggressive policy measures, as they remain stable or decline slightly since 2009.

Fourth, the weighted estimates of aggregate uncertainty obtained in this section are much less volatile than the weighted estimates presented in the previous subsections. As expected, much of the volatility in the weighted estimates shown on Figures 6 and 9 was caused by changes to the panel of survey participants. When the forecasts by panellists with the same participation patterns are used, the noise generated by entry and exit of participants is eliminated from the estimates of aggregate uncertainty.

Overall, the results presented in this section confirm that unweighted estimates of aggregate uncertainty may mislead researchers and policy-makers because they implicitly assign higher-than-optimal weights to forecasters with very poor track records. The alternative consists on using estimates of aggregate uncertainty that give more weight to the best forecast-

ers. This section has shown that these weighted estimates give rise to more reasonable and interesting dynamics of aggregate uncertainty than the unweighted estimates.

Conclusions

The SPF literature has developed many different measures of aggregate uncertainty. I advocate that all of them suffer from two problems. The first problem is what I called the *panel-composition problem*: the existing measures of aggregate uncertainty do not take into account that the panel of professional forecasters changes from one survey round to the next. Therefore, the evolution of these measures of uncertainty over time is meaningless because it compounds true changes in uncertainty with artificial changes due to the variations in the panel of respondents to the survey.

The second problem is what I called the *same-weight problem*: the existing measures of aggregate uncertainty assign the same weight to the uncertainty perceived by each survey participant, without taking forecasting performance into consideration. They do not give more weight to the information submitted by a forecaster who always performs better than the average. As a result, the unweighted estimates of aggregate uncertainty used in the literature may be biased because they may implicitly give a weight larger than optimal to underperforming forecasters.

This paper deviates from the existing literature in two dimensions. It solves the *panel-composition problem* by estimating an aggregate measure of uncertainty from the data submitted by forecasters that replied to two consecutive survey rounds. Using ECB's SPF data from 1999 Q1 to 2014 Q3, the paper finds that the effects of changes in the composition of the panel on aggregate uncertainty can be large in a statistical and economic sense. In this regard, the current standard of aggregating the results from all the participants in each survey round, independently of their participation in previous rounds, may produce very misleading results.

The paper then solves the *same-weight problem* by using performance-based weights to obtain weighted averages of individual measures of uncertainty. Forecasting performance is assessed with the logarithmic score of the density forecasts. The forecasters with better scores are assigned higher weights. The paper finds that the weighted estimates of aggregate uncertainty differ significantly from the simple averages used in the SPF literature. The differences are statistically significant and economically relevant. In particular, weighted estimates indicate a much larger increase in uncertainty than the simple averages since the start of the financial crisis. More-

over, while the unweighted estimates of aggregate uncertainty have stayed rather flat in the euro area since 2010, the weighted estimates display significant variation. The latter are much more consistent with the shocks that have hit the euro area in the last five years, like the sovereign debt crisis and the recession in 2012. They are also much more in line with the volatility displayed by the uncertainty indicators from financial markets, like the VSTOXX index.

Future research will revisit the link between uncertainty and macroeconomic outcomes using weighted estimates of aggregate uncertainty and controlling for changes in the panel of respondents to the survey. In particular, the robustness of the negative relationship between uncertainty and expected GDP growth may be explored when weighted estimates of aggregate uncertainty are used instead of simple averages.

References

- Abel, J., Rich, R., Song, J. & Tracy, J. (2015). The measurement and behavior of uncertainty: evidence from the ECB Survey of Professional Forecasters. *Journal of Applied Econometrics*. <http://dx.doi.org/10.1002/jae.2430>(Forthcoming).
- Baker, S. R., Bloom, N. & Davis, S. J. (2015). Measuring economic policy uncertainty. NBER Working Paper 21633. Cambridge, MA.
- Basu, S., & Bundick, B. (2014). Uncertainty shocks in a model of effective demand. Federal Reserve Bank of Kansas City Working Paper 14-15. <http://dx.doi.org/10.2139/ssrn.2531066>.
- Batchelor, R., & Dua, P. (1996). Empirical measures of inflation uncertainty: a cautionary note. *Applied Economics*, 28(3). <http://dx.doi.org/10.1080/000368496328704>.
- Bekaert, G., Hoerova, M. & Lo Duca, M. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7). <http://dx.doi.org/10.1016/j.jmoneco.2013.06.003>.
- Billio, M., Casarin, R., Ravazzolo, F. & van Dijk, H. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177(2). <http://dx.doi.org/10.1016/j.jeconom.2013.04.009>.
- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3). <http://dx.doi.org/10.3982/ecta6248>.
- Boero, G., Smith, J. & Wallis, K. F. (2008). Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *The Economic Journal*, 118(530). <http://dx.doi.org/10.1111/j.1468-0297.2008.02162.x>.

- Caballero, R. J. (1990). Consumption puzzles and precautionary savings. *Journal of Monetary Economics*, 25(1). [http://dx.doi.org/10.1016/0304-3932\(90\)90048-9](http://dx.doi.org/10.1016/0304-3932(90)90048-9).
- Capistrán, C., & Timmermann, A. (2009). Forecast combination with entry and exit of experts. *Journal of Business and Economic Statistics*, 27(4). <http://dx.doi.org/10.1198/jbes.2009.07211>.
- Conflitti, C. (2011). Measuring uncertainty and disagreement in the European Survey of Professional Forecasters. *Journal of Business Cycle Measurement and Analysis*, 2. <http://dx.doi.org/10.1787/jbcm-a-2011-5kg0p9zpz26k>.
- Conflitti, C., De Mol, C. & Giannone, D. (2015). Optimal combination of survey forecasts. *International Journal of Forecasting*, 31(4). <http://dx.doi.org/10.1016/j.ijforecast.2015.03.009>.
- ECB (2014). Fifteen years of the ECB Survey of Professional Forecasters. *ECB Monthly Bulletin*, January.
- Engelberg, J., Manski, C. F. & Williams, J. (2011). Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition. *Journal of Applied Econometrics*, 26(7). <http://dx.doi.org/10.1002/jae.1206>.
- García, J. A., & Manzanares, A. (2007). What can probability forecasts tell us about inflation risks? ECB Working Paper 825. Frankfurt am Main.
- Geweke, J., & Amisano, G. (2010). Optimal prediction pools. *Journal of Econometrics*, 164(1). <http://dx.doi.org/10.1016/j.jeconom.2011.02.017>.
- Genre, V., Kenny, G., Meyler, A. & Timmermann, A. (2013). Combining expert forecasts: can anything beat the simple average? *International Journal of Forecasting*, 29(1). <http://dx.doi.org/10.1016/j.ijforecast.2012.06.004>.
- Giannone, D., Henry, J., Lalik, M. & Modugno, M. (2010). An area-wide real-time database for the euro area. ECB Working Paper 1145. Frankfurt am Main.
- Gini, C. (1955). Variabilità e mutabilità. In E. Pizzeti and T. Salvemini (Eds.). *Memorie di metodologica statistica*. Rome: Libreria Eredi Virgilio Veschi.
- Giordani, P., & Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, 47(6). [http://dx.doi.org/10.1016/S0014-2921\(02\)00236-2](http://dx.doi.org/10.1016/S0014-2921(02)00236-2).
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477). <http://dx.doi.org/10.1198/016214506000001437>.
- Hall, S. G., & Mitchell, J. (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. *Oxford Bulletin of Economics and Statistics*, 67(s1). <http://dx.doi.org/10.1111/j.1468-0084.2005.00149.x>.
- Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23(1). <http://dx.doi.org/10.1016/j.ijforecast.2006.08.001>.
- Jore, A. S., Mitchell, J. & Vahey, S. P. (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 25(4). <http://dx.doi.org/10.1002/jae.1162>.
- Kascha, C., & Ravazzolo, F. (2010). Combining inflation density forecasts. *Journal of Forecasting*, 29(1-2). <http://dx.doi.org/10.1002/for.1147>.

- Kenny, G., Kostka, T. & Masera, F. (2015). Density characteristics and density forecast performance: a panel analysis. *Empirical Economics*, 48(3). <http://dx.doi.org/10.1007/s00181-014-0815-9>.
- Leahy, J. V., & Whited, T. M. (1996). The effect of uncertainty on investment: some stylized facts. *Journal of Money, Credit and Banking*, 28(1). <http://dx.doi.org/10.2307/2077967>.
- Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(70). <http://dx.doi.org/10.2307/2276207>.
- Moore, D., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2). <http://dx.doi.org/10.1037/0033-295X.115.2.502>.
- Neamtiu, M., Shroff, N., White, H. D. & Williams, C. D. (2014). The impact of ambiguity on managerial investment and cash holdings. *Journal of Business Finance & Accounting*, 41(7-8). <http://dx.doi.org/10.1111/jbfa.12079>.
- Rich, R., & Tracy, J. (2010). The relationships among expected inflation, disagreement, and uncertainty: evidence from matched point and density forecasts. *Review of Economics and Statistics*, 92(1). <http://dx.doi.org/10.1162/rest.2009.11167>.
- Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3). <http://dx.doi.org/10.1111/j.1468-0084.2008.00541.x>.
- Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6). <http://dx.doi.org/10.1002/for.928>.
- Wallis, K. F. (2006). A note on the calculation of entropy from histograms. MPRA Working Paper 52856. Munich.