Measures of Macroeconomic Uncertainty for the ECB's Survey of Professional Forecasters[1]

Víctor López-Pérez[2]

First version: 11 March 2014

This version: 3 December 2014

Abstract

This paper investigates to what extent different uncertainty measures commonly used in the SPF literature comply with a few reasonable properties. The measures published by the ECB in its quarterly report of the SPF results do not verify almost any of the properties. Unfortunately, the alternatives typically proposed in the literature do not perform much better under this metric. Instead, entropy-based measures and a new measure based on the Gini index seem more satisfactory in this regard. Independently of the measure chosen, the aggregation of the results from all the participants in each survey round may produce misleading results: they may compound true changes in uncertainty with artificial changes due to the variations in the panel of respondents to the survey. Using an aggregate measure of uncertainty from the subsets of forecasters that replied to two consecutive survey rounds, the paper finds significant increases in macroeconomic uncertainty in the euro area from 2001 to 2004, declines in uncertainty from the second half of 2004 to 2007, sharp increases from 2008 to mid-2009 and falls thereafter with the exception of the relatively more turbulent period between late 2011 and early 2012.

Keywords: uncertainty, Survey of Professional Forecasters, entropy, Gini index, European Central Bank.

JEL classification: D84, E66.

---

1. Introduction

The end of the Great-Moderation era has put the term *uncertainty* back to the front pages of newspapers and academic articles (Baker, Bloom and Davis, 2013). Uncertainty played a major role in the freeze of credit markets worldwide after the fall of Lehman Brothers in 2008.[3] It was also a key factor in the propagation or contagion of the sovereign debt crisis from Greece to other European countries in 2011.[4] Consequently, several indicators or *proxies* for uncertainty have gained prominence in economic and policy discussions.

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 and Bundick, 2012). However, the non-conventional policy actions taken by many governments and central banks in developed economies may have contributed to distort the signals extracted from financial markets in general and from proxies for uncertainty in particular (Bekaert, Hoerova and Lo Duca, 2013). Furthermore, the limited effect of these measures on "Main Street" compared to "Wall Street" may raise questions regarding the plausibility of financial market data to characterise general macroeconomic uncertainty as opposed to uncertainty in financial markets.

In this context, data from surveys may provide a more accurate picture of macroeconomic uncertainty than financial indicators. In particular, density forecasts of macroeconomic variables from professional forecasters may prove particularly valuable as they combine the expertise of highly-skilled professionals with the diversity or heterogeneity of views that naturally comes from survey methods.

The European Central Bank has used the data from its Survey of Professional Forecasters (SPF) to estimate the evolution of macroeconomic uncertainty (ECB, 2013 and 2014a, for recent examples) and the economic literature has proposed several improvements on the measures employed (e.g. Boero, Smith and Wallis, 2008). This paper takes stock of the survey based measures of uncertainty applied to SPF data in the literature and evaluates their compliance with a few "reasonable" properties.

As a preview of the results, the first finding of the paper is that the uncertainty measures published by the ECB in its quarterly report of SPF results do not verify almost any of the "reasonable" properties. Unfortunately, the alternatives typically proposed in the literature do not perform much better under this metric. Instead, entropy-based measures, relatively infrequent in the SPF literature, seem more satisfactory in this regard. The paper also includes the proposal for a new measure of macroeconomic uncertainty based on the Lorenz curve and the Gini index that verifies the "reasonable" properties as well.

Irrespectively of the measure of choice, the paper finds that the analysis of uncertainty with SPF data needs to explicitly take into account the unbalanced nature of the SPF panel of respondents. In this regard, the current standard of aggregating the results from all participants in each survey round, independently of their participation in the previous rounds, may produce very misleading results. These aggregate results may compound

---

[3] ABC news (2009): http://abcnews.go.com/Business/story?id=6559188.
[4] European Commission (2012).

true changes in uncertainty with artificial changes due to entry and exit from the panel of respondents to the survey (Engelberg, Manski and Williams, 2011). The effects of entry and exit on aggregate measures of uncertainty from the ECB's SPF are found to be sizeable and may even alter the direction of change of uncertainty measures.

Finally, the paper uses an aggregate measure of uncertainty from the data submitted by forecasters that replied to two consecutive rounds and finds significant increases in macroeconomic uncertainty in the euro area from 2001 to 2004 (the burst of the dot-com bubble), declines in uncertainty from the second half of 2004 to 2007 (the building up of the real-estate bubble), sharp increases from 2008 to mid-2009 (the start of the financial crisis), and falls thereafter with the exception of the relatively more turbulent period between late 2011 and early 2012 (the epicentre of the sovereign debt crisis).

The remainder of the paper is organised as follows. Section 2 briefly presents the ECB's SPF, describes the preliminary treatment applied to the original data provided by the ECB and defines uncertainty. Section 3 presents the properties that a "coherent" SPF-based measure of macroeconomic uncertainty should verify and evaluates to what extent measures of uncertainty commonly used in the literature comply with these properties. Section 4 discusses how SPF-based measures of uncertainty are affected by the unbalanced nature of the panel of respondents and proposes a practical solution to this problem. Section 5 summarises the results of the paper and presents directions for future research.

2. The SPF data and a brief review of the literature on SPF forecasts

2.1 The ECB's SPF dataset

The European Central Bank's Survey of Professional Forecasters has been conducted quarterly since 1999 Q1. 100 forecasters have participated at least once in the survey, although the average participation rate is around 60 forecasters per round. The panel is unbalanced, as many forecasters do not reply every quarter and some have discontinued their participation in the survey to be replaced by new panellists.

The SPF surveys point forecasts of inflation, GDP growth, unemployment, policy interest rates, compensation per employee, oil prices and exchange rates for different forecast horizons. All these variables refer to the euro area, except oil prices (Brent, in US dollars) and the exchange rate (dollar/euro). Importantly for the measurement of uncertainty, panellists are also asked to submit their density forecasts of inflation, GDP growth and unemployment by distributing probabilities among a set of predefined intervals.[5]

Therefore, the SPF provides data on the subjective probabilities individual forecasters assigned to different macroeconomic events. Each event is characterised by a macroeconomic variable, an interval and a future date. As indicated above, there are three macroeconomic variables of interest (the year-on-year inflation rate, the year–on-year GDP growth rate and the unemployment rate, all for the euro area), a time-varying set of intervals for each variable (see Figure 1) and, for most survey rounds, six forecast

---

[5] See http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html for a full description of the survey.

horizons (see Figure 2). For instance, we know that, in October 2013, forecaster number 1 assigned 70% probability to the September 2014 inflation rate in the euro area being between 1.5% and 1.9%.

How can the ECB's SPF dataset help measuring uncertainty? According to the Merrian–Webster dictionary, *uncertainty* is defined as "the quality or state of being uncertain", while *uncertain* means "not certain to occur" or "not known beyond doubt". In the context of the ECB's SPF, uncertainty could be defined as the state of a professional forecaster not knowing beyond doubt today the future value of a macroeconomic variable. The individual density forecasts from the SPF dataset may thereby be useful in this endeavour as they contain the subjective probabilities that forecasters assigned to the occurrence of different events.

But before moving on, two practical issues need to be sorted out. The first one is the selection of the forecast horizons for the construction of the uncertainty measures. Some SPF forecast horizons are constant over survey rounds (the "rolling horizons": one year ahead and two years ahead) while others are not (current calendar year, next calendar year, calendar year after the next and five calendar years ahead). Therefore, this paper focuses on the former, as uncertainty is expected to shrink mechanically if the forecast horizon shortens from one survey round to the next. Results for the five-calendar-years-ahead forecast horizon are also presented because this mechanical effect may be of lesser significance due to the length of the forecast horizon.[6]

The second practical issue relates to the probability placed in open-ended intervals. These intervals are much less informative than closed intervals for the measurement of uncertainty because of their infinite support. Previous studies have typically assumed that open-ended intervals have the same or double width than closed intervals.[7] This assumption may lead to underestimate uncertainty, especially if open-ended intervals contain relatively large probabilities. For instance, forecaster 52 in round 2009 Q1 assigned 100% probability to a one-year-ahead GDP growth rate lower than -1.0%. Given that the width of closed intervals in the ECB's SPF is 0.5%, is it reasonable to assume that she assigned 100% probability either to the [-1.5%,-1.1%] interval or to the [-2.0%,-1.1%] interval when her point forecast was -2.9%?

In order to avoid drawing wrong inferences from less informative data, any density with at least one open-ended interval that cannot *reasonably* have the same width than the closed intervals is removed from the sample. Specifically, any density with 50% more probability in an open-ended interval than in any other non-zero interval is excluded. This resulted in the exclusion of 399 inflation densities, 644 GDP-growth densities and 420 unemployment densities.[8,9,10] For the remaining densities, 6968 for inflation, 6572

---

[6] In a special questionnaire on forecasting practices and techniques conducted by the ECB among SPF panellists in 2013, most respondents indicated that their five-calendar-years-ahead forecasts of real GDP growth and unemployment can be interpreted as their estimates for long-term potential output growth (68%) and the non-accelerating inflation rate of unemployment (53%) respectively. See http://www.ecb.europa.eu/stats/prices/indic/forecast/shared/files/resultssecondspecialquestionnaireecbsurveyspf201401en.pdf?92f09eaa6906c6e15bf77fb4dcdd055f for the results of this special questionnaire.

[7] See, for instance, Batchelor and Dua (1996).

[8] For completeness, densities with 100% probability in open-ended intervals are also excluded.

[9] Densities that verified this criterion but have less than 1% probability in open-ended intervals are not removed from the sample. Otherwise, all computer-generated densities with support from -∞ to ∞ (e.g. a normal density function) would be excluded.

for GDP growth and 6315 for unemployment, open-ended intervals are treated as having the same width than closed intervals. Thus, the next sections do not make any distinctions between closed and open-ended intervals: all intervals are assumed to be closed and assumed to have the same width.

2.2 Brief survey of the literature on SPF forecasts

Abundant research has been conducted using the SPF *point* forecasts, especially for the US-SPF maintained by the Federal Reserve Bank of Philadelphia because its time series are longer than those from the ECB's SPF. For instance, Branch and Evans (2005) used point forecasts to show that simple constant-gain learning algorithms fit well the forecasts submitted by US-SPF panellists. Rubaszek and Skrzypcznski (2008) found that most of the US-SPF forecasts of GDP, inflation and the short-term interest rate are more accurate than those from small DSGE and VAR models, but Wieland and Wolters (2011) stated that model forecasts compare particularly well to professional forecasts during the economic recoveries. Manzan (2011) advocated that the heterogeneity in the point forecasts comes from the different interpretation by different forecasters of the same news, in particular at longer horizons.

Point forecasts were also used by Capistran and Timmermann (2009a) to explore the bias found in US-SPF inflation expectations and claimed that it cannot be explained either by an asymmetric loss or rational expectations. Similarly, Harvey and Newbold (2003) pointed out that forecasts errors from the US-SPF have non-zero mean and are not normally distributed Croushore (2010) claimed, however, that the systematic bias disappears when real-time data is considered, and Wang and Lee (2014) recently stated that forecast rationality under asymmetric loss fails to be rejected for most of the rolling periods they use. Pierdzioch, Ruelke and Stadtmann (2010) found that oil-price SPF forecasts are biased away from the consensus forecasts (anti-herding behaviour).

Individual SPF forecasts have also triggered a discussion on the best way to combine them into an aggregate SPF forecast. Poncela, Rodriguez and Sanchez-Mangas (2011) argued that combination methods based on partial least squares, principal component regressions and factor analyses may perform better that the usual average of the individual US-SPF forecasts. However, Genre, Kenny, Meyler and Timmermann (2013) used the ECB's SPF forecasts to show that the simple equally-weighted average of the professional forecasts is rarely outperformed by other combination methods.

Point expectations from the SPF have been used to test theoretical forward-looking models. For instance, Smith (2009), Adam and Padula (2011) and López-Pérez (2014) used point forecasts from the US-SPF and the ECB's SPF respectively to estimate the parameters of the New Keynesian Phillips Curve. Paloviita and Viren (2014) found that individual SPF forecasters seem to behave according to an uncertainty-augmented hybrid specification of the New Keynesian Phillips Curve and that inflation uncertainty has a negative impact on economic activity by lowering output growth, boosting inflation and reducing the price-sensitiveness of aggregate supply.

---

[10] Five other densities were also removed because their probabilities do not sum up to 100, even after rounding (all of them for inflation). Furthermore, the probabilities of other five inflation densities, five GDP densities and one unemployment density were rounded to sum 100.

Research using the *density* forecasts from the SPF is also gathering speed. Engelberg, Manski and Williams (2009) found that the US-SPF point forecasts are well approximated by the central tendencies of the forecasters' subjective distributions, but Clements (2010, 2011) argued that these two sets of forecasts are sometimes inconsistent because i) some forecasters do not update their density forecasts when new information arrives and ii) the forecasters typically round their probability distributions. Galbraith and van Norden (2011) and Lahiri and Wang (2013) suggested that US-SPF density forecasts are incorrectly calibrated. Giordani and Soederlind (2006), Soederlind (2009) and Chua, Kim and Suardi (2011) found evidence of overconfidence in the US-SPF density forecasts, while Clements (2014) recently argued that there may be signs of underconfidence at forecast horizons shorter than a year. Kenny, Kostka and Masera (2014) have also found overconfidence in the density forecasts from the ECB's SPF.

Finally, some authors have constructed measures of uncertainty from the density forecasts submitted by the SPF panellists (see, among others, Giordani and Soederlind, 2003, Boero, Smith and Wallis, 2008, Engelberg, Manski and Williams, 2009 and 2011, Rich and Tracy, 2010, and Rich, Song and Tracy, 2012). Their contributions are described and assessed in the Section 3.2.

3. "Coherent" measures of uncertainty for the ECB's SPF

This section has been inspired by the initial developments of the so-called "mathematically-coherent risk measures" in the field of finance. Artzner, Delbaen, Eber and Heath (1999) presented and justified a set of four "desirable properties for measures of risk and call[ed] the measures satisfying these properties *coherent*" (p. 203). After presenting their four axioms, they analysed the extent to which prominent measures of financial risk (the SPAN margin system by the Chicago Mercantile Exchange, the Securities and Exchange Commission rules used by the National Association of Security Dealers, and the variance-quantile method of Value-at-Risk) complied with these properties. More recent research has extended the "mathematically-coherent risk measures" to multi-period settings (Fritelli and Scandolo, 2006), financial market equilibria (Heath and Ku, 2004), higher-order moments (Krokhmal, 2007) and to positive deviations from expected returns (Chen and Wang, 2008). Gotoh, Shinozaki and Takeda (2013) developed robust methods to estimate and minimise the "coherent" measures of risk.

A similar approach is used in this paper. Unfortunately, the four properties presented by Arztner *et al.* (1999) are not directly applicable to the SPF uncertainty measures because the former were designed to identify financial investments that do not belong to an "acceptance set", i.e. a financial position with an unacceptable future net worth. For instance, their approach identifies as a risk a portfolio choice that yields a negative net worth equal to, let's say, -1 *with 100% probability*. While this is certainly useful from a prudential regulatory perspective or from a risk-management perspective, it is less so for the analysis of SPF uncertainty measures. Consequently, a different set of axioms or properties are needed here.

The properties used in this paper are instead borrowed from the field of information theory, initiated by Shannon (1948). Information theory deals with the engineering problem of the transmission of information over a noisy channel. To this end, this field

has developed measures of the amount of information revealed by the occurrence of events (e.g. Kolmogorov, 1968), the best known of them probably being Shannon's entropy (Shannon, *op. cit.*). In his seminal paper, Shannon listed a number of "reasonable" properties that a good measure of information should have. Then, he showed how the entropy of an information source complies with these "reasonable" properties and advocated its use for the measurement of information.

These measures of information have already been applied to the analysis of economic issues: for instance, Theil (1967) used Shannon's entropy to construct measures of income inequality. They can also be used to measure uncertainty from a probability distribution. As Uffink (1990) affirmed, in the context of the problem of designing measures of uncertainty from density forecasts or probability distributions,

> "a closely related situation occurs in economics or sociology when one studies, for example, the distribution of wealth in a certain society or population. In that case it is interesting to ask whether this wealth is more or less evenly distributed over the individual members of the society, or on the contrary, largely concentrated in the hands of a 'happy few'. Thus one finds oneself facing the problem of finding a mathematical expression which measures the 'concentration' or 'inequality' in a given distribution. This problem is mathematically completely analogous to our problem, although, if the distributed quantity is wealth rather than probability, one will not interpret such an expression in terms of predictability."

As a matter of fact, Rich and Tracy (2010) applied Shannon's entropy to compute measures of uncertainty for the US SPF. Therefore, it seems logical to borrow some of the "reasonable" properties enunciated by Shannon[11] and examine to what extent measures of uncertainty commonly used in the SPF literature verify them.

3.1 Properties of a "coherent" measure of macroeconomic uncertainty for the ECB's SPF

A measure of uncertainty for the SPF is a mapping function, U, from $\mathcal{R}^n$ to $\mathcal{R}$, where n is the number of intervals available to SPF forecasters for a variable of interest. It translates a set of n probabilities, one for each interval, into a real number, the measure of uncertainty. This subsection lists and justifies the reasonable properties that a measure of uncertainty for the ECB's SPF should comply with. A measure of uncertainty that verifies the following properties is said to be "coherent".

The first property is related to the value of the uncertainty measure when the SPF forecaster that submits the probability distribution of an expected variable is certain about its future value. It seems reasonable that, in the absence of uncertainty, a "coherent" measure of uncertainty takes its minimum value. As pointed out in the previous section, SPF panellists are asked to submit their density forecasts of inflation, GDP growth and unemployment by distributing probabilities among a set of predefined intervals. Therefore, a SPF forecaster that is certain about the future value of the forecasted variable would assign 100% probability to the interval containing the

---

[11] Some of the properties listed by Shannon are not applicable to SPF data as they refer to the joint probability of several events.

expected future value of the variable. Consequently, it seems reasonable that a "coherent" measure of uncertainty should take its minimum value when the density forecast of the variable of interest is degenerated.

Property 1: A "coherent" measure of uncertainty takes its minimum value if and only if all the probability is assigned to just one interval.[12]

$$U(p_{l1}, p_{l2},..., p_{ln}) = \min(U(p_1, p_2,..., p_n)) \quad iff \quad p_{li} = 100 \quad with \quad 1 \le i \le n \quad and \quad p_{lj} = 0$$
$$\forall \ j \ne i.$$

where p_k denotes the probability allocated by a forecaster to interval k.

Let's take the example of forecaster 56 in round 1999 Q3 (July), who assigned 100% probability to a June 2001 inflation rate between 1.5% and 1.9%. As we do not have information regarding how probabilities are distributed within each interval, a satisfactory measure of uncertainty should take its minimum value when evaluating a forecast like this.

It may happen that a forecaster that splits 100% probability between two adjacent intervals feels less uncertainty than forecaster 56 in our example. This may be the case if the range of the density forecast is very narrow but includes the bound of an interval, forcing the forecaster to split the probability mass. The SPF data, however, does not allow the identification of these cases.

The second property is related to the value of the uncertainty measure in the hypothetical case in which the SPF forecaster that submitted the density forecast of a variable of interest had no information at all about its future value. It seems reasonable that, under complete uncertainty, a "coherent" measure of uncertainty takes its maximum value. How would a SPF forecaster distribute probabilities among the different intervals when he has absolutely no idea about the future value of the forecasted variable? As all intervals in the SPF have the same width, a forecaster that does not have any information about the future value of a variable should distribute his 100% probability allocation uniformly across all intervals. This is so because, under complete uncertainty, the forecaster is unable to discriminate between different intervals: they are all the same to him and, thereby, he would assign the same probability to all of them. Consequently, it seems reasonable that a "coherent" measure of uncertainty should take its maximum value when the same probability is allocated to every interval.

Property 2: A "coherent" measure of uncertainty takes its maximum value if and only if the same probability is allocated to every interval.[13]

$$U(p_{l1}, p_{l2},..., p_{ln}) = \max(U(p_1, p_2,..., p_n)) \ iff \ p_{li} = p_{lj} \ \forall \ i,j.$$

---

[12] This is property 1 that makes the entropy a "reasonable measure of choice or information" in Shannon (1948).

[13] This is property 2 that makes the entropy a "reasonable measure of choice or information" in Shannon (1948).

The third property requires that a "coherent" uncertainty measure monotonically decreases from its maximum to its minimum. Let's think again about the forecaster with no information at all about the future value of a variable. By Property 2, this forecaster would assign the same probability to every SPF interval. Now, let's assume that this forecaster starts to gather useful information about the future value of the variable. He could then reduce the probability placed in the intervals that are less compatible with the new information and increase the probability assigned to the intervals that are more coherent with the new information. In this case, it seems reasonable that a "coherent" measure of uncertainty should decrease. If the forecaster continued to gather useful information about the future value of the variable of interest, he would assign more probability to fewer intervals. As a result, it seams reasonable that a "coherent" measure of uncertainty should fall more and more. Finally, let's assume that this forecaster were so successful in obtaining new useful information about the future that he is now certain about the future value of the variable of interest. By Property 1, he would assign 100% probability to the interval containing the expected future value of the variable. Then, a "coherent" measure of uncertainty would reach its minimum.

Property 3: _If, as a result of the transfer of some probability between two intervals, the probabilities assigned to these intervals become less (more) equal, a "coherent" measure of uncertainty falls (increases)._[14]

$$U(p_1,...,p_i+\varepsilon,...,p_j-\varepsilon,...,p_n) < U(p_1,...,p_i,...,p_j,...,p_n) \quad if \quad p_j - p_e \leq 0 \quad for \quad any$$
$1 \leq i, j \leq n$, and $\varepsilon > 0$.
$$U(p_1,...,p_i+\varepsilon,...,p_j-\varepsilon,...,p_n) > U(p_1,...,p_i,...,p_j,...,p_n) \quad if \quad p_j - p_i > \varepsilon \quad for \quad any$$
$1 \leq i, j \leq n$, and $\varepsilon > 0$.

For instance, Figure 3 shows the individual density forecasts of the unemployment rate one year ahead submitted by forecasters 103 and 96 in the 2013 Q4 SPF round. The density by forecaster 96 equals the density by forecaster 103 plus a transfer of 5% probability from interval 11.5-11.9% to interval 10.5-10.9%. As the former interval contains much more probability than the latter, Property 3 requires that a "coherent" measure of uncertainty should assign higher uncertainty to the density submitted by forecaster 96 and lower uncertainty to the density sent by forecaster 107.

The next, and last, property is not borrowed from Shannon (1948). It does not refer to the measure of uncertainty itself but to its aggregation mechanism. This aggregation mechanism maps the individual values of the uncertainty measure, which are calculated from density forecasts submitted by individual forecasters, into an aggregate measure of uncertainty for a variable of interest. Therefore, the aggregation mechanism is a mapping function from $\mathcal{R}^F$ to $\mathcal{R}$, where F is the number of individual forecasters who submitted density forecasts of a variable of interest. An example of aggregation mechanism is the average of the values of the uncertainty measure obtained from the individual density forecasts of the inflation rate one year ahead: it combines the outcomes of applying the uncertainty measure to each individual density forecast of the inflation rate one year ahead into an aggregate measure of uncertainty for expected inflation one year ahead.

---

[14] This is property 4 that makes the entropy a "reasonable measure of choice or information" in Shannon (1948).

What does this property of a "coherent" aggregation mechanism require? Let's assume that two forecasters submitted their density forecasts of a variable of interest to the ECB. Let's also assume that, from each individual density forecast, a value of the measure of uncertainty is obtained, U_1 and U_2. Finally, let's assume that the first forecaster is very certain about the future and the value of U_1 is relatively low, let's say $U_1 = 1$. The second forecaster feels much less sure about the future and the value of U_2 is relatively high, let's say $U_2 = 10$. What should the value of the aggregate measure of uncertainty be for this pair of forecasters? It seems reasonable that it is somewhere between 1 and 10. Neither as sure as the confident forecaster nor as uncertain as the doubtful forecaster, but somewhere in between. This is what this property requires from a "coherent" aggregation mechanism.

Property AM (aggregation mechanism): Let's assume k+1 forecasters with values of the uncertainty measure for a variable of interest equal to U_1, U_2,…,U_k and U_{k+1}, with $k \geq 1$. A "coherent" aggregation mechanism, AM, of the values of the uncertainty measure verifies that:

$$AM(U_1,U_2,...,U_k,U_{k+1}) < AM(U_1,U_2,...,U_k) \text{ if } U_{k+1} < AM(U_1,U_2,...,U_k) \text{ and}$$
$$AM(U_1,U_2,...,U_k,U_{k+1}) > AM(U_1,U_2,...,U_k) \text{ if } U_{k+1} > AM(U_1,U_2,...,U_k) \text{ and}$$
$$AM(U_1,U_2,...,U_k,U_{k+1}) = AM(U_1,U_2,...,U_k) \text{ if } U_{k+1} = AM(U_1,U_2,...,U_k)$$

This property ensures that, if a forecaster with low individual uncertainty enters a group of forecasters with higher aggregate uncertainty, the aggregate uncertainty of the new group is lower than the aggregate uncertainty of the old group. For instance, Figure 4 shows the density forecasts of the inflation rate five calendar years ahead submitted by forecasters 94, 98 and 107 in the 2013 Q3 SPF round. These panellists did not change their density forecasts in the next round (2013 Q4). Panellist 48, who did not reply in 2013 Q3, submitted a density forecast in 2013 Q4 (Figure 5). If a measure of uncertainty assigns a lower value to forecaster 48 compared to the group of forecasters 94, 98 and 107, a "coherent" aggregation mechanism should yield a lower value of aggregate uncertainty for the group {48, 94, 98, 107} compared to the group {94, 98, 107}.

3.2. Uncertainty measures for the SPF: do they satisfy the four properties?

This subsection reviews the measures of uncertainty most frequently used with SPF data and explores whether they are "coherent". To the extent that the properties listed in the previous subsection are found to be reasonable, they may be useful to provide an additional angle, not present yet in the literature, to discriminate between the existing measures of uncertainty that verify these criteria and those that do not.

3.2.1 *Disagreement*

Disagreement is defined as the standard deviation of individual point forecasts. This measure is published by the European Central Bank in its quarterly SPF report[15] and has been suggested as a good proxy for uncertainty when density forecasts are not available (Giordani and Soederlind, 2003, Wallis, 2004).[16]

Figure 6 displays the nine disagreement time series calculated from the SPF data (three variables times three horizons).[17] The declining trend in disagreement is interrupted by the start of the financial crisis in 2008. Disagreement spikes in 2009 and tends to fall afterwards (with the exception of unemployment five calendar years ahead).

The definition of disagreement makes it particularly vulnerable to outliers. For instance, the spike in disagreement for inflation two years ahead in 2003 Q2 is caused by forecaster 73, who submitted a point forecast of -1% while no other forecaster submitted a figure below 1%. Interestingly, panellist 73's point forecasts in 2003 Q1 and 2003 Q3 were 1.2% and 1.3% respectively, which may lead to think that the 2003 Q2 figure was a mistake, but the density forecast submitted in 2003 Q2 is consistent with a negative point forecast. With the aim of limiting the vulnerability to outliers, some researchers have suggested the use of the quasi-standard deviation, i.e. half the distance between the 16th and the 84th percentiles of the distribution of point forecasts, to compute disagreement (Giordani and Soederlind, 2003).[18].

Disagreement does not take into account the information contained in density forecasts. Therefore, it is not a mapping function from $\mathfrak{R}^n$ to $\mathfrak{R}$, where n is the number of intervals available to SPF forecasters for a variable of interest. Disagreement does not translate a set of n probabilities, one for each interval, into a real number, and thereby it is not a measure of uncertainty. Consequently, it cannot comply with Properties 1 to 3 and cannot be "coherent".[19] As disagreement cannot be computed at the individual level by construction, Property AM does not apply.

3.2.2 *Moment-based measures*

Probably the most frequently-used measures of uncertainty are based on the standard deviation of a probability distribution, either an individual density forecast or an aggregation of them. Prominent examples are the standard deviation of the aggregate probability distribution (featuring in each quarterly ECB report on SPF results and in ECB, 2014a), the mean standard deviation of the individual density forecasts (Giordani and Soederlind, 2003), the root mean subjective variance or RMSV (Batchelor and Dua, 1996) and the implied RMSV (Boero, Smith and Wallis, 2008).

Figures 7 to 10 show the result of computing these uncertainty measures with the SPF data. For the calculation of a standard deviation from SPF density forecasts, either aggregate or individual, an assumption on how the probability is distributed inside each

---

[15] For inflation forecasts five calendar years ahead only.

[16] For evidence against the use of disagreement as a proxy for uncertainty see Lahiri and Liu (2006), Rich and Tracy (2010) and Lahiri and Sheng (2010).

[17] Every point forecast is included in the calculation of disagreement.

[18] This suggestion will be followed when the implied measures of uncertainty are computed below.

[19] The same applies to the measures of uncertainty based on the ex-post forecast errors (see, for instance, Hayford and Malliaris, 2012).

interval is needed. For the measures shown in these charts, it is assumed that all the probability allocated to each interval is concentrated in the middle point of the interval. We will come back to the relevance of this assumption later on.

Although there is a general pattern in the time series behaviour of these series (falling trend in uncertainty before the crisis, significant increase at the start of the crisis and mild decline afterwards) significant differences remain among these statistics. The most striking are probably:

- The surge in the standard deviation of the aggregate probability distribution of GDP growth one year ahead in 2009, which is not as pronounced in other measures of uncertainty. This is the mechanic consequence of skyrocketing disagreement in point forecasts, as the variance of the aggregate probability distribution can be decomposed in the average variance of the individual density forecasts and the variance of the distribution of point forecasts (Wallis, 2005):[20]

$$\sigma_A^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \frac{1}{n}\sum_{i=1}^{n}(f_i - \frac{1}{n}\sum_{i=1}^{n}f_i)^2 \qquad [1]$$

where σ_A^2 is the variance of the aggregate probability distribution, σ_i^2 denotes the variance of the density forecast submitted by panellist i and f_i is the point forecast by panellist i.

- An implied RMSV for unemployment two years ahead equal to zero. Actually, it is not zero but an imaginary number with real part equal to zero. RMSV applies equation [1] to obtain the average individual variance from the average aggregate variance and the variance of the point forecasts.[21,22] Unfortunately, in 2009 Q2 the variance of the point forecasts of the unemployment rate two years ahead reaches its maximum (1.21), higher than the variance of the aggregate probability distribution (1.16). The square root of a negative number yields an imaginary number.[23]

Do these moment-based measures satisfy the properties described in Section 3.1? Are they "coherent" measures of uncertainty? The standard deviation reaches its lowest value when all the probability is assigned to just one interval, thus Property 1 is satisfied. Property 2, however, is not because the maximum standard deviation is obtained when 50% probability is allocated to the lowest interval and the remaining 50% to the highest interval, leaving the intervals in between empty.[24]

---

[20] Only under the assumption that the individual point forecasts are the means of the individual density forecasts.

[21] This indirect procedure avoids the use of individual density forecasts which, in the context of the ECB's SPF, are probably too narrow to be consistent with rationality (Kenny, Kostka and Masera, 2014).

[22] For the calculation of the variance of point forecasts, the point forecasts submitted by panellists whose density forecasts have not been excluded in Section 2 are used. If a panellist submitted a density forecast but not a point forecast, the mean of the density forecast is used instead. As pointed out in footnote 18, the *quasi-standard deviation* is used.

[23] This isolated episode happens somewhat more frequently under alternative assumptions about the distribution of the probability inside each interval (see Figure 11).

[24] There remain 87 bimodal densities in the clean SPF database.

Property 3 is not satisfied either by the standard deviation as a measure of uncertainty: the standard deviation increases when probability is transferred from intervals closer to the mean of the distribution to intervals further away from the mean, irrespective of how much probability these intervals contained.

Apart from not being "coherent", the standard deviation as a measure of uncertainty suffers from a technical problem: as pointed out above, the calculation of the mean and the standard deviation from a histogram requires some assumptions regarding how the probability is distributed within each interval. Some studies assume that all the probability allocated to an interval is in the middle point of each interval[25] (Rich and Tracy, 2010) while others fit a continuous distribution to the histogram. The normal distribution (Boero, Smith and Wallis, 2012) and the generalised beta distribution (Engelberg, Manski and Williams, 2011, and Rich, Song and Tracy, 2012) are the most commonly used while the piecewise linear distribution (Conflitti, 2011) and the skewnormal distribution (García and Manzanares, 2007) have also been proposed.

How robust is the standard deviation to these assumptions? Figure 11 depicts different time series of the implied RMSV statistic under four different assumptions: *midpoint*, *normal*, *beta* and *modified midpoint*. The *midpoint* line is the result of assuming that all the probability of the aggregate probability distribution is located in the middle point of each interval. *Normal* fits a normal distribution to the aggregate probability distribution.[26] *Beta* fits a generalised beta distribution instead.[25] Finally, *mod midpoint* is a slight modification of *midpoint*: it assumes that all the probability allocated to one interval is assigned to just one single point, but not necessarily to the middle point. For any given interval and SPF round, if the distance between the middle point of the interval and the average point forecast in that round is at least 0.1 percentage points, the middle point is not used. Instead, all the probability allocated to that interval is moved 0.1 percentage points closer to the average point forecast from the middle point of the interval.[27] For instance, the average inflation point forecast one year ahead was 1.5% in 2013 Q4. For the calculation of the implied RMSV statistic for inflation one year ahead in round 2013 Q4, all the probability allocated to interval [2.0%, 2.4%] will not be assigned to its middle point, 2.2%, but to 2.2% - 0.1% = 2.1%.

Not surprisingly, the *midpoint* assumption yields higher levels of uncertainty than the others with the *normal*, *beta* and *mod midpoint* alternatives clustering together more often than not.[28] The *midpoint* and the *mod midpoint* assumptions lead to almost identical variations in uncertainty. Although, for most of the time, the direction of change is the same across all alterative assumptions, differences remain. For example, the *normal* assumption indicates a decline in uncertainty for inflation five calendar years ahead in 2001 Q2 but the other assumptions suggest increases in uncertainty; for GDP growth five calendar years ahead, uncertainty declined according to the *midpoint* and *mod midpoint* assumption in 2008 Q2, it increased modestly under the *beta* assumption

---

[25] An alternative is to assume that the probability is spread uniformly across each interval (Rich, Song and Tracy, 2012), which automatically increases the variance by one-twelfth of the squared bin width (Boero, Smith and Wallis, 2008).

[26] If the histogram has less than three intervals with positive probability, a triangular approximation is fitted instead (Engelberg, Manski and Williams, 2009).

[27] If the difference between the middle point of the interval and the average point forecast is less than 0.1 percentage points, the middle point of the interval is used.

[28] Values of zero on the charts represent imaginary numbers with real part equal to zero, which appear when the implied mean subjective variance takes a negative value.

and rose much more under the *normal* assumption; for unemployment one year ahead, uncertainty rose according to the *midpoint* and *mod midpoint* assumptions in 2012 Q3 but measures based on the other two alternative assumptions experienced a decline.[29] Therefore, the standard deviation as a measure of uncertainty lacks robustness with regard to the assumptions made on the allocation of probabilities inside each interval.

As regards Property AM (the aggregation mechanism), the average individual standard deviation verifies it because the mean of any individual measure of uncertainty is a "coherent" aggregation mechanism. This is not the case for the variance of the aggregate probability distribution used by the ECB in its quarterly report of SPF results, because the individual density forecasts are added up interval by interval before calculating the variance of the resulting aggregate distribution. To illustrate this, see Figure 4 and 5 again. Forecaster 48 submitted in 2013 Q4 a density forecast for inflation five calendar years ahead with variance 0.06 (mid-point assumption), lower than the variance of the aggregate distribution of the group of forecasters 94, 98 and 107 (0.59). However, the variance of the aggregate distribution of the group of forecasters 48, 94, 98 and 107 is 0.64, higher than 0.59. Therefore, the variance of the aggregate probability distribution does not comply with Property AM and is not a "coherent" aggregation mechanism.

3.2.3 *The interquartile range (IQR)*

Used, among others, by Engelberg, Manski and Williams (2011), it is defined as the median distance between the 25th and 75th quantiles of the individual density forecasts. Rich, Song and Tracy (2012) claimed that its main advantage with respect to moment-based measures is its robustness to outliers since it uses the median instead of the mean as aggregation mechanism. Figure 12 shows the IQR computed with the ECB's SPF data, which provides a qualitatively similar picture of the evolution of uncertainty compared to the previous measures: mild decline up to 2007 followed by an increase until 2010 and a stabilisation at relatively high levels afterwards.[30]

The IQR verifies Property 1 as the minimum range is obtained when all the probability falls in just one interval. It does not comply with Property 2, however, since the maximum IQR is obtained with density forecasts that do not place the same probability in each interval, but half the probability in each of the extreme intervals.

Property 3 is not always verified, because the transfers of probability between intervals within the bounds of the IQR do not change the IQR. The same happens to the transfers of probability between intervals to the left (right) of the 25th (75th) quantile. As a result, the IQR is not a "coherent" measure of uncertainty either.

As in the case of moment-based measures, the calculation of the IQR requires the use of assumptions to distribute probabilities inside each interval. As shown above, in the context of measures based on the standard deviation, these assumptions are not

---

[29] Very similar results are obtained when the RMSV is computed from individual density forecasts instead of from equation [1], with the obvious exception of the imaginary numbers. These results are available from the author upon request.

[30] Following Rich, Song and Tracy (2012), the probability allocated to an interval is assumed to be distributed uniformly inside it.

innocuous: different assumptions give rise to different uncertainty figures for the same underlying dataset of density forecasts. Even the direction of change in the uncertainty measure may not be robust to alternative assumptions.

Finally, the election of the median as aggregation mechanism, justified by the robustness to outliers, breaks Property AM: the median of a series does not always change when a number below (above) the median is added to it. This may happen, for instance, if several individual density forecasts share the same IQR. Therefore, the median of individual measures of uncertainty is not a "coherent" aggregation mechanism.

3.2.4 *Entropy*

As pointed out above, the entropy is a concept borrowed from the field of information theory. It represents the value of new information (Shannon, 1948). When uncertainty is high, new information is highly valuable and entropy is also high. When there is no uncertainty, new information adds no value and entropy is minimal. Rich and Tracy (2010) were, to the best of my knowledge, the first that used entropy-based measures of uncertainty for the US SPF. For the purpose of measuring uncertainty from the ECB's SPF histograms, entropy may be calculated as follows (Wallis, 2006):

$$e = \log w - \sum_{j=1}^{n} p_j \log p_j \qquad [2]$$

where the first term on the right-hand side is the log of the interval width, w, which is constant in the ECB's SPF, n is the number of intervals in the histogram and p_j is the probability (between 0 and 1) allocated to interval j.[31] The second term on the right-hand side is the relevant term for uncertainty measurement and its average individual value is plotted in Figure 13. It shows the already familiar pattern of a modest decline in uncertainty before the financial crisis, a relatively abrupt increase in 2008-2009 and a stabilisation at high levels, or even a modest rise, until the end of the sample.

Is the average individual entropy a "coherent" measure of uncertainty? It is by definition because the properties of a "coherent" measure of uncertainty were borrowed from the reasonable properties of the entropy found by Shannon (1948). Property 1 is verified because the second term in [2] is minimised when all but one probabilities are zero. This can be easily seen in Figure 14, which shows a simple example with only two intervals and two probabilities, p and 1-p.[32] The entropy takes its lowest value when 100% probability is assigned to just one interval (Property 1) and its highest value when the probability is evenly distributed across intervals (Property 2). It is also trivial to check compliance with Property 3 as the entropy is monotonically increasing from its minimum to its maximum when probability is transferred from the most likely interval to the less likely one.

---

[31] If any $p_j = 0$, it is assumed that $p_j \log p_j = 0$.

[32] Only the second term of the entropy formula is shown.

Finally, the average individual entropy verifies Property AM because it makes use of the arithmetic mean as the aggregation mechanism of individual entropies. The average individual entropy is thus a "coherent" measure of uncertainty.

3.2.5. *The Gini index*

This paper presents a new measure of uncertainty for the ECB's SPF inspired by the application by Theil of Shannon's entropy to the measurement of income inequality. The Gini index (Gini, 1955) is a concept borrowed from the literature on income and wealth inequality and is based on the Lorenz curve (Lorenz, 1905). This curve is typically used to represent how much wealth is in the hands of the poorest *x*% of the population. The Lorenz curve may also be applied to the analysis of uncertainty with SPF data by representing the cumulative probability allocated to the *x%* less likely intervals. If a density forecast were characterised by minimum uncertainty (with 100% probability placed in one interval), the Lorenz curve would be zero before jumping to 100% in the last interval. On the contrary, if a density forecast represents maximum uncertainty (the same probability allocated to every interval), the Lorenz curve would increase in regular steps from the first interval to the last.

Figure 15 exemplifies the use of Lorenz curves with the SPF data. The density forecast of inflation five calendar years ahead submitted by forecaster 42 in 2007 Q3 exhibits only three intervals with positive probability, with 70% probability in the [1.5, 1.9%] interval. This relatively low level of uncertainty pushes the Lorenz curve close to the bottom-right corner on the chart. Two years later, in 2009 Q3, forecaster 16 submitted a density forecast for the same variable and forecast horizon but placed positive probabilities in all but one of the 14 intervals provided by the ECB. The highest probability placed in a single interval by this forecaster was 16% (in the [2.0, 2.4%] interval). This relatively higher level of uncertainty yields a Lorenz curve much closer to the 45-degree line, which represents the maximum level of uncertainty (same probability in every interval).

A caveat is needed here because the attentive reader may have noted that the ECB only provided 9 intervals for the inflation density forecasts in 2007 Q3 (Figure 1), while there are 14 values on the *x*-axis of Figure 15. The nature of the 'data-cleaning' procedure described in Section 2, and in particular the removal of all the densities with suspiciously large probabilities in the open intervals, makes it plausible the addition of empty intervals to both sides of the interval range when the maximum number of intervals (14) is not provided to the forecasters. In the context of Figure 15, forecaster 42 is assumed to have assigned 0% probability in 2007 Q3 to the intervals [-2.5%, -2.1%], [-2.0%, -1.6%], [-1.5%, -1.1%], [-1.0%, -0.6%] and [+4.0%, +4.4%], which seems reasonable given that the nearest intervals were also empty. Such assumptions are not needed for forecaster 16 because all the 14 intervals were available in the 2009 Q3 round. The alternative approach, in which the empty intervals are not added for the computation of the Lorenz curves, suffers from the lack of comparability between the Gini indices for the periods with a different number of intervals. Given the relatively high frequency of changes in the number of the intervals available to the forecasters (see Figure 1 again), this alternative approach was discarded.

From the Lorenz curves obtained from each individual density forecast, the calculation of the individual Gini indices is straightforward. The Gini index is defined as the distance between the 45-degree line and the Lorenz curve divided by the area below the 45-degree line:

$$G = -\frac{\sum_{i=1}^{n}(x_i - lc_i)}{\sum_{i=1}^{n} x_i}$$

where n is the number of intervals, x is the $nx1$ vector of ordinates representing the 45-degree line, $(1/n, 2/n,\ldots, 1)$', and lc is the $nx1$ vector of ordinates from the Lorenz curve. As the original Gini index declines with uncertainty, the index has been multiplied by -1 for an easier comparability with the other uncertainty measures. Figure 16 shows the average (negative) Gini indices for inflation, GDP growth and unemployment computed with the SPF data. The same pattern emerges, with a decline in uncertainty before the rise that took place in 2008.

Is the average Gini index a "coherent" measure of uncertainty for the SPF? The Gini index verifies Property 1, because its minimum value (zero) is only reached when 100% probability is allocated to just one interval. It also complies with Property 2 as its maximum value (one) is only obtained when the Lorenz curve coincides with the 45-degree line that denotes maximum uncertainty. Property 3 holds as well because transfers from high-probability intervals to low-probability intervals raise a section of the Lorenz curve leaving the other sections unchanged. Finally, as already indicated above, the arithmetic average of individual measures of uncertainty complies with Property AM. Hence, the average Gini index is a "coherent" measure of uncertainty for the SPF.

3.3. Assessment

This section has investigated to what extent different uncertainty measures commonly used in the SPF literature are "coherent" measures of uncertainty, i.e. they comply with a few reasonable properties borrowed from the field of information theory. The measures published by the ECB in its quarterly report of SPF results do not verify almost any of the properties. Unfortunately, the alternatives proposed in the literature do not perform much better under this metric. Instead, entropy-based measures, used by Rich and Tracy (2010), and a new measure based on the Gini index seem more satisfactory in this regard.[33]

4. Evaluating changes in the time series of macroeconomic uncertainty from the SPF

---

[33] This statement does not mean that the other measures of uncertainty must be discarded. As Cramér (1946) pointed out, "each measure has advantages and disadvantages of its own, and a measure that renders excellent service in one case may be more or less useless in another".

Once the time series of an aggregate measure of macroeconomic uncertainty is obtained from a survey in general, and from the ECB's SPF in particular, its analysis typically takes the form of a visual inspection (ECB quarterly report on SPF results[34]) or to the calculation of confidence bands around it (Boero, Smith and Wallis, 2008). These analyses however do not address the problem of changes in the composition of the panel of the survey.

Engelberg, Manski and Williams (2011), in their analysis of the US SPF, recommended to go beyond aggregate figures and to examine the changes to the individual responses of the survey. This is the avenue this section takes, trying to ascertain the effects on the aggregate measures of uncertainty caused by the entry and exit of forecasters from the SPF panel.

Figure 17 shows the percentage changes in (i) the aggregate measure of uncertainty using the complete pool of respondents[35] and in (ii) the aggregate measure of uncertainty using the subset of panellists that submitted individual density forecasts during two consecutive rounds. The Gini index is used as the uncertainty measure of choice.[36]

Two conclusions can be extracted from the charts. First, the effects on the variability of the uncertainty measure caused by changes in the composition of the panel may be sizeable. For instance, the panel "Inflation five calendar years ahead" shows that most of the largest changes in the aggregate uncertainty measure are mainly a consequence of entry and exit, while variations in an aggregate measure of uncertainty that does not include entries and exits are much more muted.

The second conclusion is that not taking into account entry and exit from the panel may significantly affect the results, not only in terms of the size of the effects, but even on the direction of changes.[37] For instance, an analyst that looks at the whole pool of forecasters may conclude that the review of the ECB's monetary strategy, whose results were published in May 2003 and in which the ECB's inflation objective was clarified[38], led to an *increase* of uncertainty regarding long-term inflation rates (see panel "Inflation five calendar years ahead", 2003 Q3).[39] But if the analyst looks at the difference in the forecasts submitted by panellists that replied to both the 2003 Q2 and 2003 Q3 surveys, a mild *decrease* in uncertainty would be found. A second example is the perceived *decline* in the uncertainty surrounding GDP growth rates one year ahead in 2011 Q4 and 2012 Q1 when an aggregate measure of uncertainty averaging across all respondents is used. This result, in fact is just the consequence of exit and entry: when exits and entries are excluded, uncertainty actually *increased* during these quarters.

---

[34] ECB (2014b): "Aggregate uncertainty surrounding longer-term inflation expectations, as measured by the standard deviation of the aggregate probability distribution, eased slightly but remains around the relatively high level observed since 2009."

[35] SPF data filtered as indicated in Section 2.

[36] The results are qualitatively similar when the average individual entropy or the average standard deviation of individual density forecasts is chosen as the measure of uncertainty. Results are available from the author upon request.

[37] See Capistran and Timmermann (2009b). They proposed an algorithm for back-filling missing observations to balance the US-SPF panel.

[38] ECB (2003).

[39] The 2003 Q2 survey was conducted in April, before the results of the review were published in June.

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 0.45 (GDP growth five years ahead) to 0.73 (GDP growth one year ahead). Therefore, the use of a measure of uncertainty that aggregates the results from all the respondents may lead to mistakes. Instead, a focus on an aggregate measure of uncertainty that only looks at the subset of forecasters that replied to two consecutive rounds may produce more accurate results.

In this context, if the random sampling assumption is relaxed,[40] the Wilcoxon signed-rank test (Wilcoxon, 1945, and Siegel, 1956) may identify whether the changes in the aggregate measures of uncertainty are statistically significant.[41] Figure 18 shows the results of the test for the differences in the average Gini indices of uncertainty from one quarter to the next.[42] A red (green) bar denotes a statistically significant increase (decrease) in average uncertainty with respect to the previous quarter. While there are only a few statistically significant changes in uncertainty with respect to the previous quarter, the appearance of several consecutive (non-significant) grey bars in the same direction hints to statistically significant changes in uncertainty if more distant reference periods were chosen (see, for instance, the panel "Inflation five calendar years ahead" during the period 2008 Q2 – 2009 Q1).

Figure 19 puts the nine panels shown on Figure 18 together in order to get a sense of the overall changes in macroeconomic uncertainty from the ECB's SPF. The picture reveals significant increases in uncertainty from 2001 to 2004, declines in uncertainty from the second half of 2004 to 2007, sharp increases from 2008 to mid-2009 and falls thereafter with the exception of the relatively more turbulent period between late 2011 and early 2012.

5. Conclusion

This paper has reviewed the main measures of macroeconomic uncertainty applied to the SPF data. It evaluates the compliance of each of these measures with four properties that may be regarded as "reasonable". The first finding of the paper is that the uncertainty measures published by the ECB in its quarterly report of the SPF results do not verify almost any of the properties. Unfortunately, the alternatives typically proposed in the literature do not perform much better under this metric. Instead, entropy-based measures, relatively infrequent in the SPF literature seem more satisfactory in this regard. The paper also includes a new measure of macroeconomic uncertainty based on the Lorenz curve and the Gini index that verifies the "reasonable" properties as well.

---

[40] As in Boero, Smith and Wallis (2008). The sample of respondents to the ECB's SPF is not a random sample extracted from the population of professional forecasters in the European Union.

[41] The test compares pairs of values (individual measures of uncertainty) for the same set of individuals (forecasters) in two different periods (survey rounds). Therefore, only the forecasters that replied to those two survey rounds are considered, dodging the problems caused by entry and exit.

[42] The results are qualitatively similar when the average individual entropy or the average standard deviation of individual density forecasts is chosen as the measure of uncertainty. Results are available from the author upon request.

Irrespectively of the measure chosen for the analysis of uncertainty, the paper finds that the unbalanced nature of the SPF panel needs to be accounted for in the analysis of SPF data. In this regard, the aggregation of the results from all the participants in each survey round, independently of their participation in the previous rounds, may produce very misleading results. They may compound true changes in uncertainty with artificial changes due to entry and exit from the panel of respondents to the survey. The effects of entry and exit on the aggregate measures of uncertainty are found to be sizeable and may even alter the direction of change of the uncertainty measure. Therefore, the comparison of the aggregated results from the subset of forecasters that participated in each of several survey rounds may need to become the standard procedure for the analysis of uncertainty from the SPF data.

By proceeding in this way, i.e. by using an aggregate measure of uncertainty from the subsets of forecasters that replied to two consecutive rounds, the paper founds significant increases in macroeconomic uncertainty in the euro area from 2001 to 2004 (the burst of the dot-com bubble), declines in uncertainty from the second half of 2004 to 2007 (the build-up of the real-estate bubble), sharp increases from 2008 to mid-2009 (the start of the financial crisis), and falls thereafter with the exception of the relatively more turbulent period between late 2011 and early 2012 (the epicentre of the sovereign debt crisis).

Further research is expected to explore the link between changes in individual uncertainty measures from SPF density forecasts and individual point forecasts. The aim of that project would be to better understand the link between uncertainty and macroeconomic outcomes, e.g. if higher uncertainty is associated with lower GDP growth forecasts, higher inflation forecasts and higher unemployment forecasts at the individual level.

References

Adam, K., and M. Padula (2011), "Inflation Dynamics and Subjective Expectations in the United States", *Economic Inquiry* 49, pp. 13-25.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999), "Coherent Measures of Risks", *Mathematical Finance* 9, pp. 203-228.

Baker, S.R., N. Bloom and S.J. Davis (2013), "Measuring Economic Policy Uncertainty", Chicago Booth Research Paper 13-02.

Basu, S., and B. Bundick (2012), "Uncertainty Shocks in a Model of Effective Demand", Boston College Working Paper 774.

Batchelor, R.A., and P. Dua (1996), "Empirical Measures of Inflation Uncertainty: a Cautionary Note", *Applied Economics* 28, pp. 333-341.

Bekaert, G., M. Hoerova and M. Lo Duca (2013), "Risk, Uncertainty and Monetary Policy", *Journal of Monetary Economics* 60, pp. 771-788.

Berger, H., M. Erhmann and M. Fratzscher (2009), "Forecasting ECB Monetary Policy: Accuracy is a Matter of Geography", *European Economic Review* 53, pp. 1028-1041.

Boero, G., J. Smith and K.F. Wallis (2008), "Uncertainty and Disagreement in Economic Prediction: the Bank of England Survey of External Forecasters", *The Economic Journal* 118, pp. 1107-1127.

Boero, G., J. Smith and K.F. Wallis (2012), "The Measurement and Characteristics of Professional Forecasters' Uncertainty", submitted to the *Journal of Applied Econometrics* and available here: http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/boero/bsw8_jae_final.pdf

Branch, WA, and G.W. Evans (2005), "A Simple Recursive Forecasting Model", *Economic Letters* 91, pp. 158-166.

Capistran, C., and A. Timmermann (2009a), "Disagreement and Biases in Inflation Expectations", *Journal of Money, Credit and Banking* 41, pp. 365-396.

Capistran, C., and A. Timmermann (2009b), "Forecast Combination with Entry and Exit of Experts", *Journal of Business & Economic Statistics* 27, pp. 428-440.

Chen, Z.P., and Y. Wang (2008), "Two-Sided Coherent Risk Measures and their Application in Realistic Portfolio Optimization", *Journal of Banking & Finance* 32, pp. 2667-2673.

Chua, C.L., D. Kim and S. Suardi (2011), "Are Empirical Measures of Macroeconomic Uncertainty Alike?", *Journal of Economic Surveys* 25, 801-827.

Clements, M.P. (2010), "Explanations of the Inconsistencies in Survey Respondents' Forecasts", *European Economic Review* 54, pp. 536-549.

Clements, M.P. (2011), "An Empirical Investigation on the Effects of Rounding on the SPF Probabilities of Decline and Output Growth Histograms", *Journal of Money, Credit and Banking* 43, pp. 207-220.

Clements, M.P. (2014), "Forecast Uncertainty-Ex Ante and Ex Post: US Inflation and Output Growth", *Journal of Business & Economic Statistics* 32, pp. 206-216.

Conflitti, C. (2011), "Measuring Uncertainty and Disagreement in the European Survey of Professional Forecasters", *Journal of Business Cycle Measurement and Analysis* 7, pp. 69-103.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press.

Croushore, D. (2010), "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data", *B.E. Journal of Macroeconomics* 10, pp. 1-32.

ECB (2003), "The Outcome of the ECB's Evaluation of its Monetary Policy Strategy", in the June issue of the ECB Monthly Bulletin, pp. 79-92.

ECB (2013), "Results of the ECB Survey of Professional Forecasters for the Fourth Quarter of 2013", Box 7 in the November issue of the ECB Monthly Bulletin, pp. 61-65.

ECB (2014a), "Fifteen Years of the ECB Survey of Professional Forecasters", in the January issue of the ECB Monthly Bulletin, pp. 55-67.

ECB (2014b), "Results of the ECB Survey of Professional Forecasters for the First Quarter of 2014", Box 8 in the February issue of the ECB Monthly Bulletin, pp. 60-65.

Engelberg, J., C.F. Manski and J. Williams (2009), "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters", *Journal of Business and Economic Statistics* 27, pp. 30-41.

Engelberg, J., C.F. Manski and J. Williams (2011), "Assessing the Temporal Variation of Macroeconomic Forecasts by a Panel of Changing Composition", *Journal of Applied Econometrics* 26, pp. 1059-1078.

European Commission (2012), "Risk and Uncertainty in Euro Area Sovereign Debt Markets and their Impact on Economic Activity", in the 2012-04 Quarterly Report on the Euro Area, pp. 31-38.

Fritelli, M., and G. Scandolo (2006), "Risk Measures and Capital Requirements for Processes", *Mathematical Finance* 16, pp. 589-612.

Galbraith, J.W., and S. van Norden (2011), "Kernel-Based Calibration Diagnostics for Recession and Inflation Probability Forecasts", *International Journal of Forecasting* 27, pp. 1041-1057.

Garcia, J.A., and A. Manzanares (2007), "What Can Probability Forecasts Tell Us about

Inflation Risks?", ECB Working Paper 825.

Genre, V., G. Kenny, A. Meyler, A. Timmermann (2013), "Combining Expert Forecasts: Can Anything Beat the Simple Average?", *International Journal of Foecasting* 29, pp. 108-121.

Gini, C. (1955), "Variabilità e mutabilità", in E. Pizzeti and T. Salvemini (eds.) *Memorie di metodologica statistica*, Rome.

Giordani, P., and P. Soederlind (2003), "Inflation Forecast Uncertainty", *European Economic Review* 47, pp. 1037–59.

Giordani, P., and P. Soederlind (2006), "Is There Evidence of Pessimism and Doubt in Subjective Distributions? Implications for the Equity Premium Puzzle", *Journal of Economic Dynamics & Control* 30, pp. 1027-1043.

Gotoh, J.Y., K. Shinozaki and A. Takeda (2013), "Robust Portfolio Techniques for Mitigating the Fragility of CVaR Minimization and Generalization to Coherent Risk Measures", *Quantitative Finance* 13, pp. 1621-1635.

Harvey, D.I., and P. Diebold (2003), "The Non-Normality of Some Macroeconomic Forecast Errors", *International Journal of Forecasting* 19, pp. 635-653.

Hayford, M.D., and A.G. Malliaris (2012), "Transparent US Monetary Policy: Theory and Tests", *Applied Economics* 44, pp. 813-824.

Heath, D., and H.J. Ku (2004), "Pareto Equilibria with Coherent Measures of Risk", *Mathematical Finance* 14, pp. 163-172.

Kenny, G., T. Kostka and F. Masera (2014), "How Informative Are the Subjective Density Forecasts of Macroeconomists?", *Journal of Forecasting* 33, pp. 163-185.

Kolmogorov, A. (1968), "Three Approaches to the Quantitative Definition of Information", *International Journal of Computer Mathematics* 2, pp. 157-168.

Krokhmal, P.A. (2007), "Higher Moment Coherent Risk Measures", *Quantitative Finance* 7, pp. 373-387.

Lahiri, K., and F. Liu (2006), "Modelling Multi-period Inflation Uncertainty Using a Panel of Density Forecasts", *Journal of Applied Econometrics* 21, pp. 1199-1219.

Lahiri, K., and X.G. Sheng (2010), "Measuring Forecast Uncertainty by Disagreement: the Missing Link", *Journal of Applied Econometrics* 25, pp. 514-538.

Lahiri, K., and J.G. Wang (2013), Evaluating Probability Forecasts for GDP Declines Using Alternative Methodologies", *International Journal of Forecasting* 29, pp. 175-190.

López-Pérez, V. (2014), "Do Professional Forecasters Behave as if They Believed in the New Keynesian Phillips Curve for the Euro Area", submitted to the ECB Working Paper Series.

Lorenz, M.O. (1905), "Methods of Measuring the Concentration of Wealth", *Publications of the American Statistical Association* 9, pp. 209–219.

Manzan, S. (2011), "Differential Interpretation in the Survey of Professional Forecasters", *Journal of money, Credit and Banking* 43, pp. 993-1017.

Paloviita, M., and M. Viren (2014), "Inflation and Output Growth Uncertainty in Individual Survey Expectations", *Empirica* 41, pp. 69-81.

Pierdzioch, C., J.C. Ruelke and G. Stadtmann (2010), "New Evidence of Anti-Herding of Oil-Price Forecasts", *Energy Economics* 32, pp. 1456-1459.

Poncela, P., J. Rodriguez and R. Sanchez-Mangas (2011), "Forecast Combination Through Dimension Reduction Techniques", *International Journal of Forecasting* 27, pp. 224-237.

Rich, R., J. Song and J. Tracy (2012), "The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters", Federal Reserve Bank of New York Staff Report 588.

Rich, R., and J. Tracy (2010), "The Relationships among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts", *Review of Economics and Statistics* 92, pp. 200-207.

Rubaszek, M., and P. Skrzypcznski (2008), "On the Forecasting Performance of a Small-Scale DSGE Model", *International Journal of Forecasting* 24, pp. 498-512.

Shannon, C.E. (1948), "The Mathematical Theory of Communication", *The Bell System Technical Journal* 27, pp. 379–423, 623–656.

Siegel, S. (1956), *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Series in Psychology.

Smith, G.W. (2009), "Pooling Forecasts in Linear Rational Expectations Models", *Journal of Economic Dynamics and Control* 33, pp. 1858-1866.

Soederlind, P. (2009), "The C-CAPM with Ex Post Data", *Journal of Macroeconomics* 31, pp. 721-729.

Theil, H. (1967), "The Measurement of Income Inequality", in *Economics and Information Theory*, Rand McNally & Company, Chicago. Chapter 4.

Uffink J.B.M. (1990), "Measures of Uncertainty and the Uncertainty Principle", University of Utrech, *PhD* Thesis available in: http://www.projects.science.uu.nl/igg/jos/publications/proefschrift.pdf

Wallis, K.F. (2004), "Forecast Uncertainty, its Representation and Evaluation", tutorial lectures, IMS Singapore, 3-6 May.

Wallis, K.F. (2005), "Combining Density and Interval Forecasts: A Modest Proposal", *Oxford Bulletin of Economics and Statistics* 67, pp. 983-994.

Wallis, K.F. (2006), "A Note on the Calculation of Entropy from Histograms", MPRA Paper 52856.

Wang, Y.Y., and T.H. Lee (2014), "Asymmetric loss in the Greenbook and the Survey of Professional Forecasters", *International Journal of Forecasting* 30, pp. 235-245.

Wieland, V., and M.H. Wolters (2011), "The Diversity of Forecasts from Macroeconomic Models of the US Economy", *Economic Theory* 47, pp. 247-292.

Wilcoxon, F. (1945), "Individual Comparisons by Ranking Methods", *Biometrics Bulletin* 1, pp. 80–83.

Tables and Figures

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

Variable	Forecast horizon	Correlation coefficient (1999:2-2013:4)
Inflation	1 year ahead	0.68
	2 years ahead	0.66
	5 calendar years ahead	0.53
	All	0.60
GDP growth	1 year ahead	0.73
	2 years ahead	0.61
	5 calendar years ahead	0.45
	All	0.59
Unemployment	1 year ahead	0.63
	2 years ahead	0.53
	5 calendar years ahead	0.64
	All	0.60
All	All	0.60

Figure 1: Intervals available in the ECB's SPF
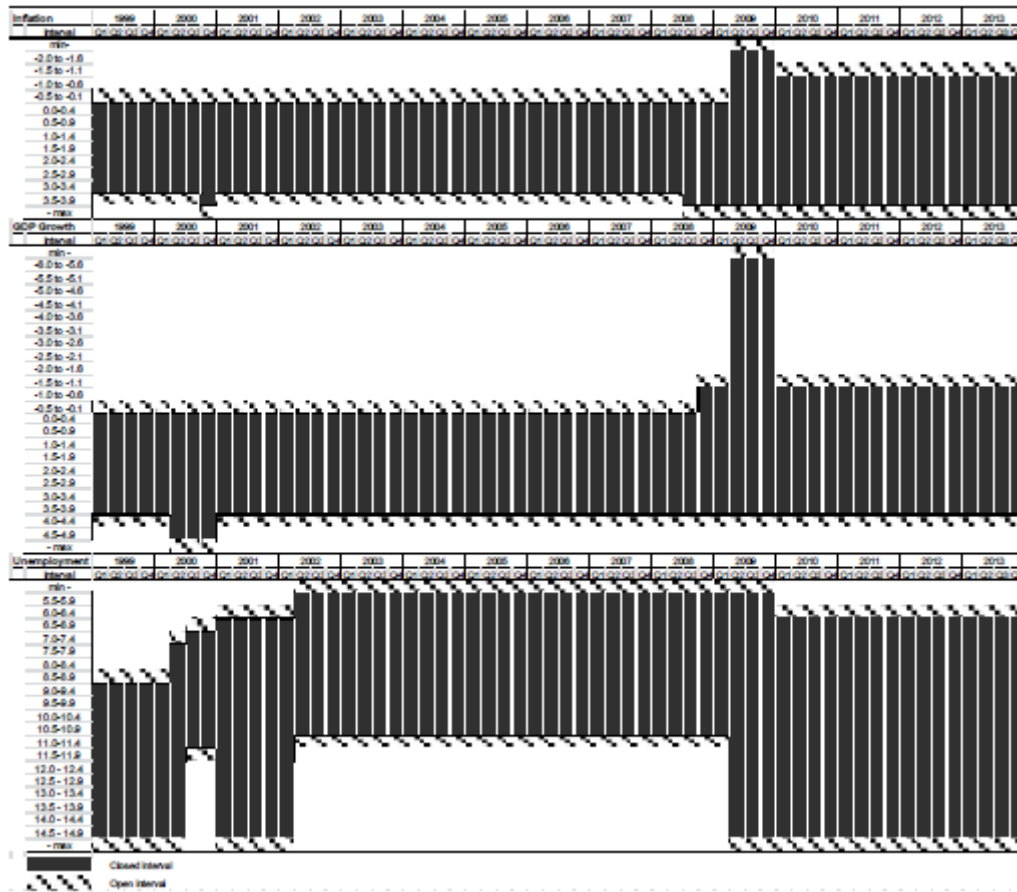
Source: ECB.

Figure 2: Forecast horizons in the ECB's SPF

forecast horizons surveyed in that particular round

Source: ECB.

Figure 3: Selected one-year-ahead unemployment density forecasts from the 2013 Q4 SPF round
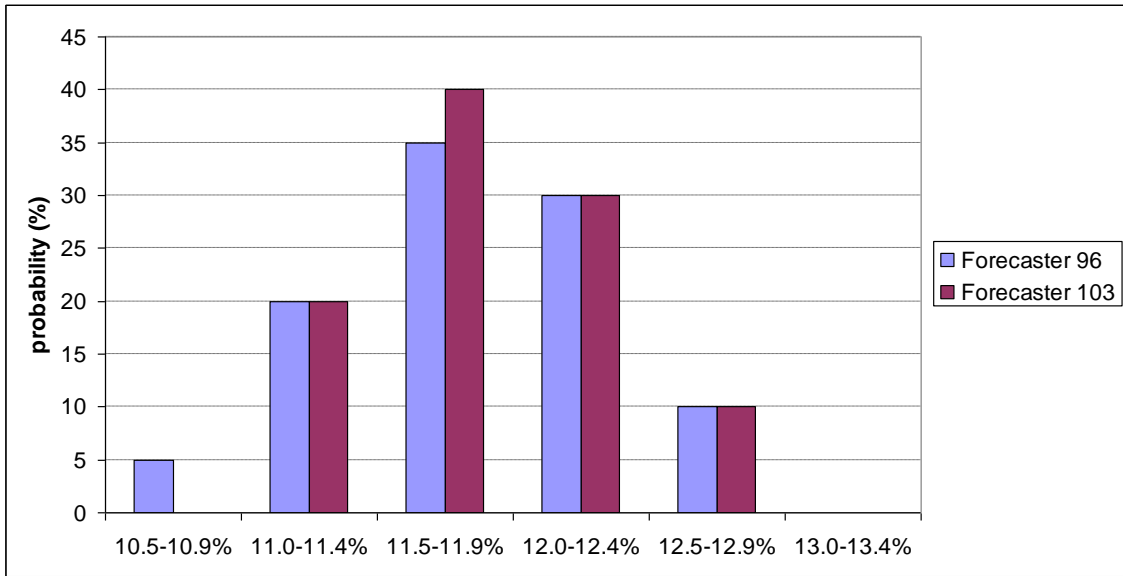
Figure 4: Selected inflation density forecasts five calendar years ahead from the 2013 Q3 SPF round
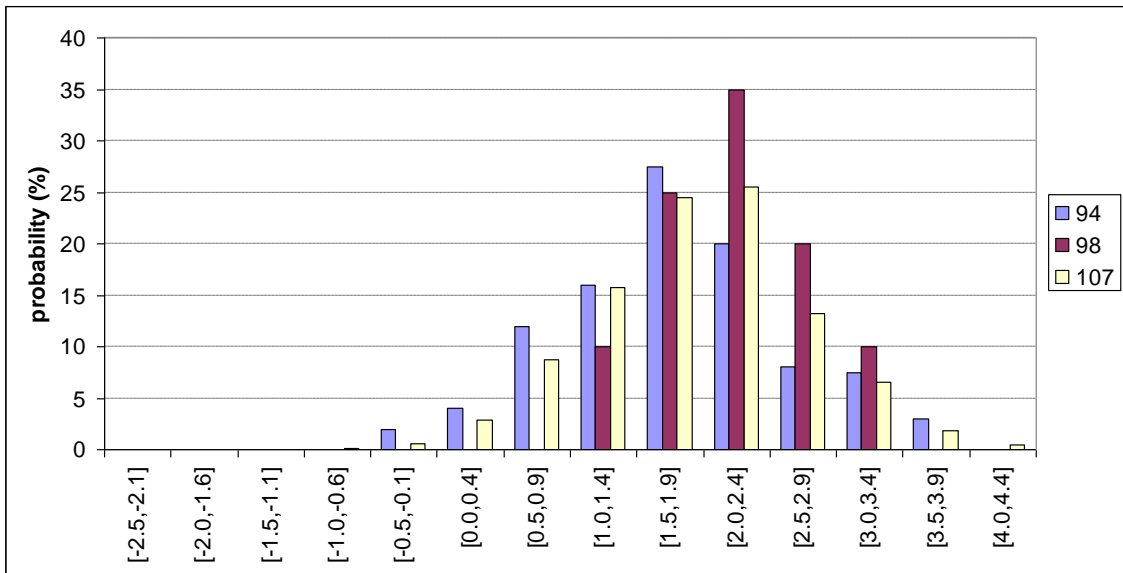
Figure 5: Selected inflation density forecasts five calendar years ahead from the 2013 Q4 SPF round
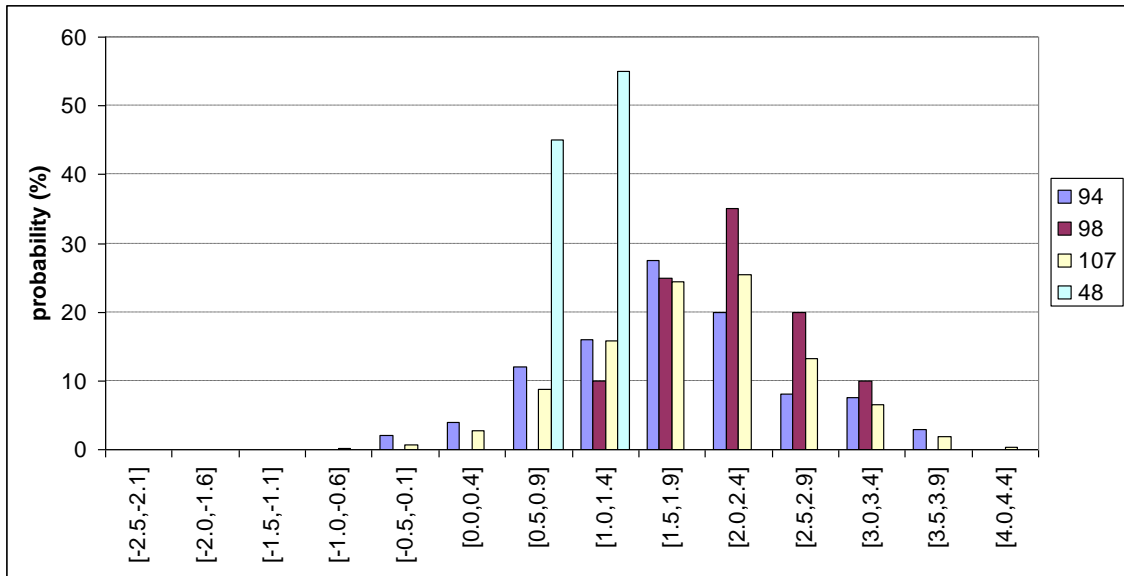
Figure 6: Disagreement in the ECB's SPF

Inflation

GDP growth

Unemployment
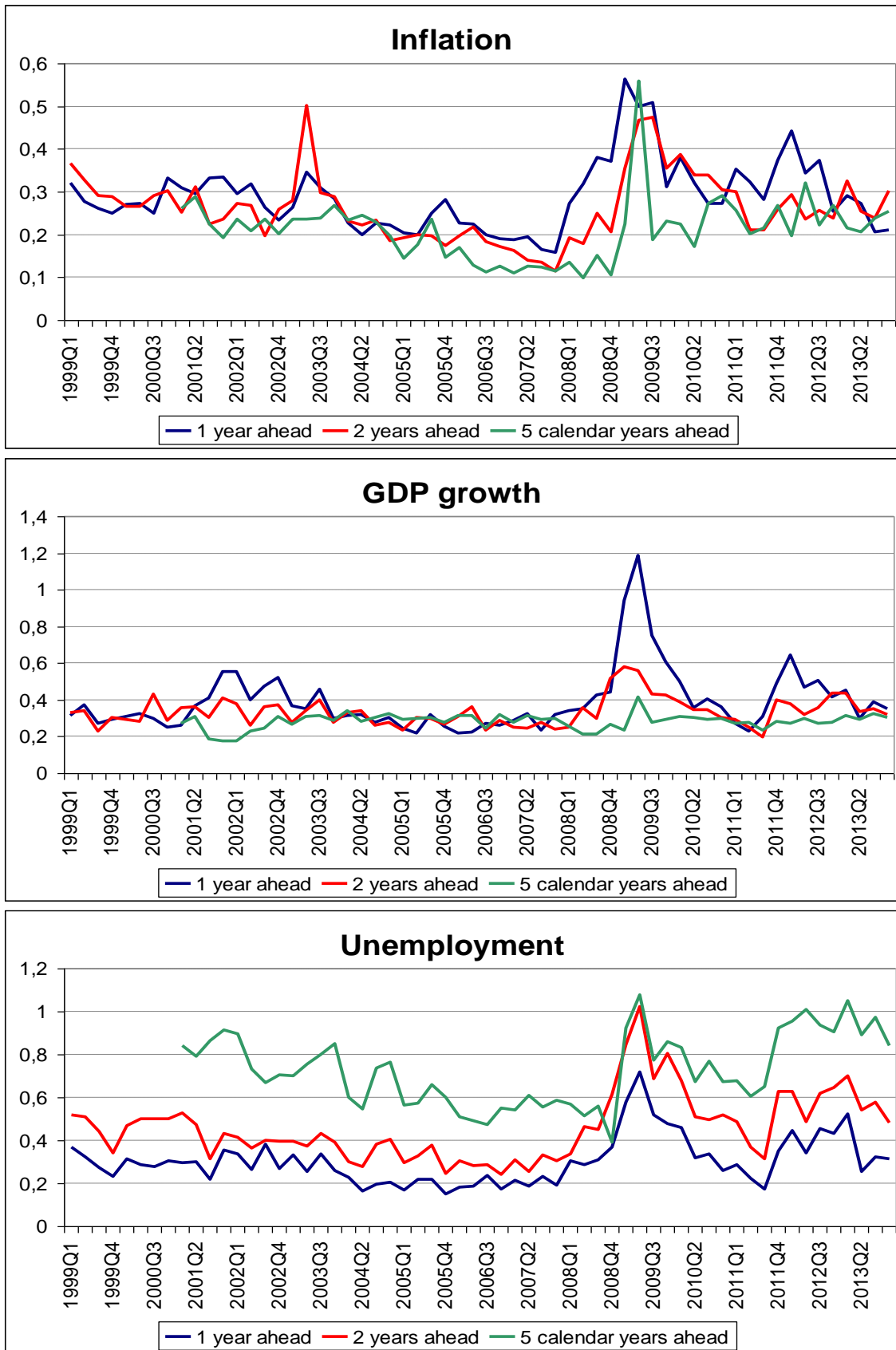
Figure 7: Standard deviations of aggregate probability distributions from the SPF

Inflation

1 year ahead — 2 years ahead — 5 calendar years ahead

GDP growth

1 year ahead — 2 years ahead — 5 calendar years ahead

Unemployment

1 year ahead — 2 years ahead — 5 calendar years ahead
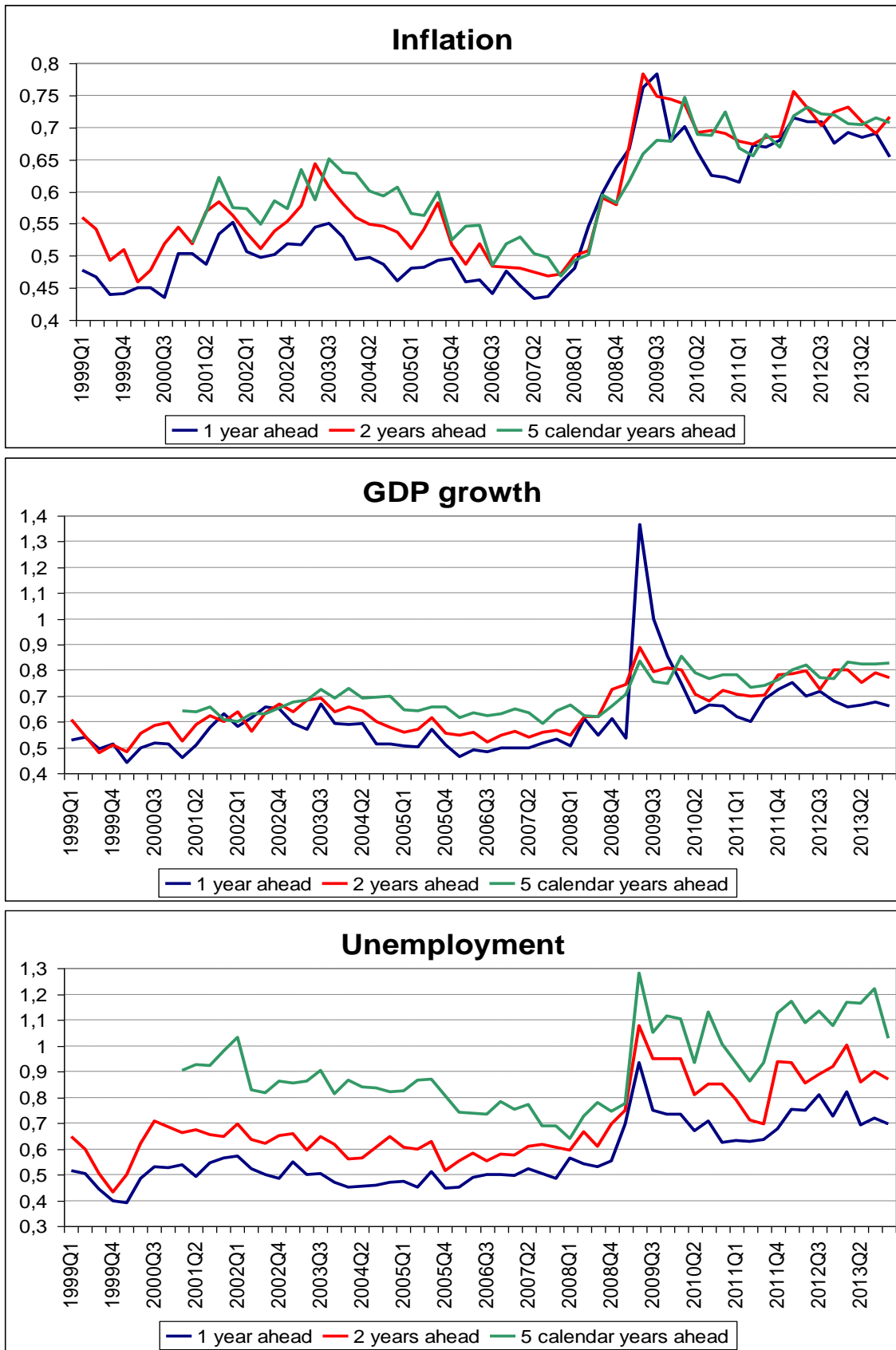
Figure 8: Average standard deviations of the individual density forecasts
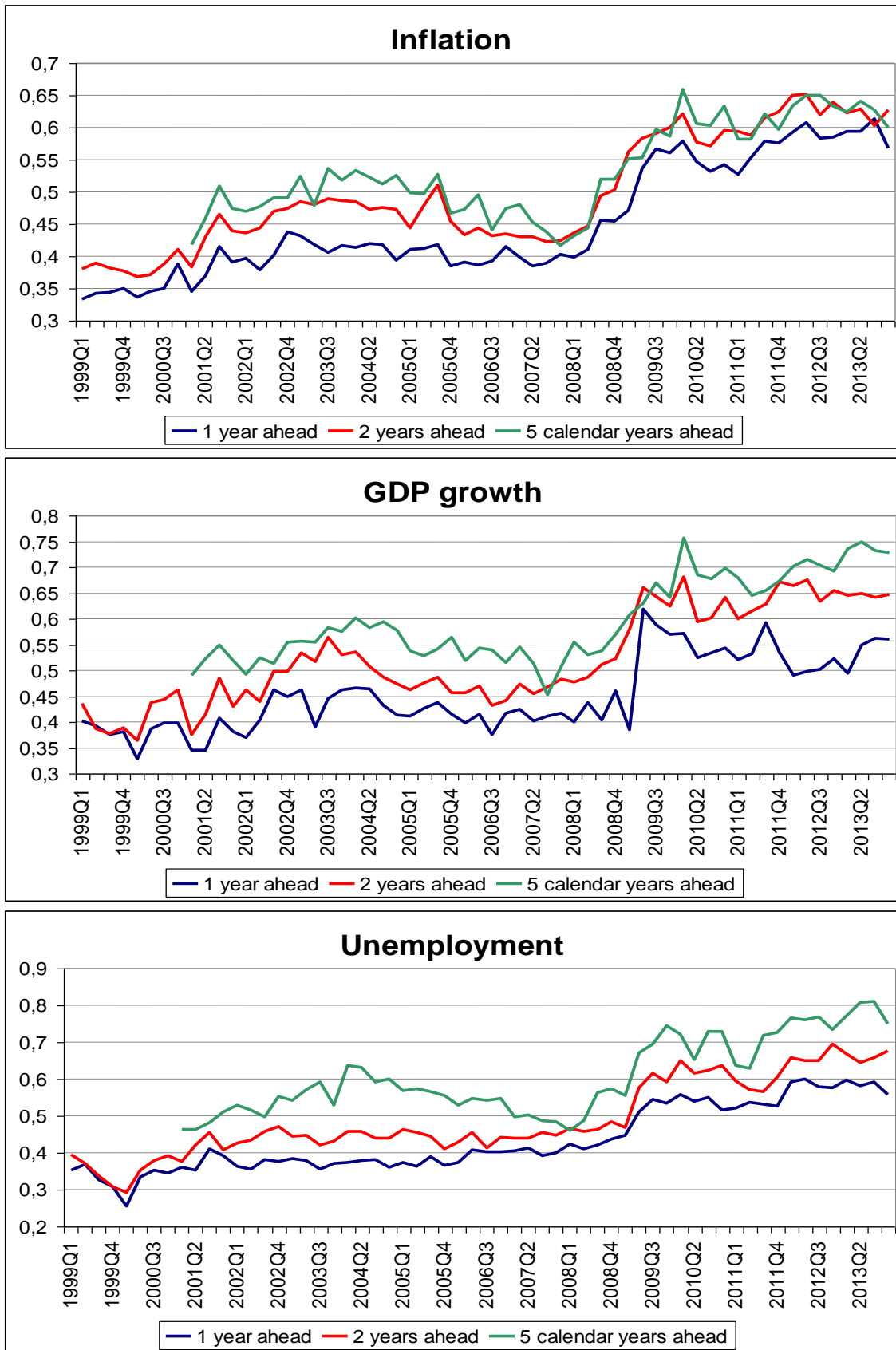
Figure 9: RMSV of the individual density forecasts

Inflation

GDP growth

Unemployment
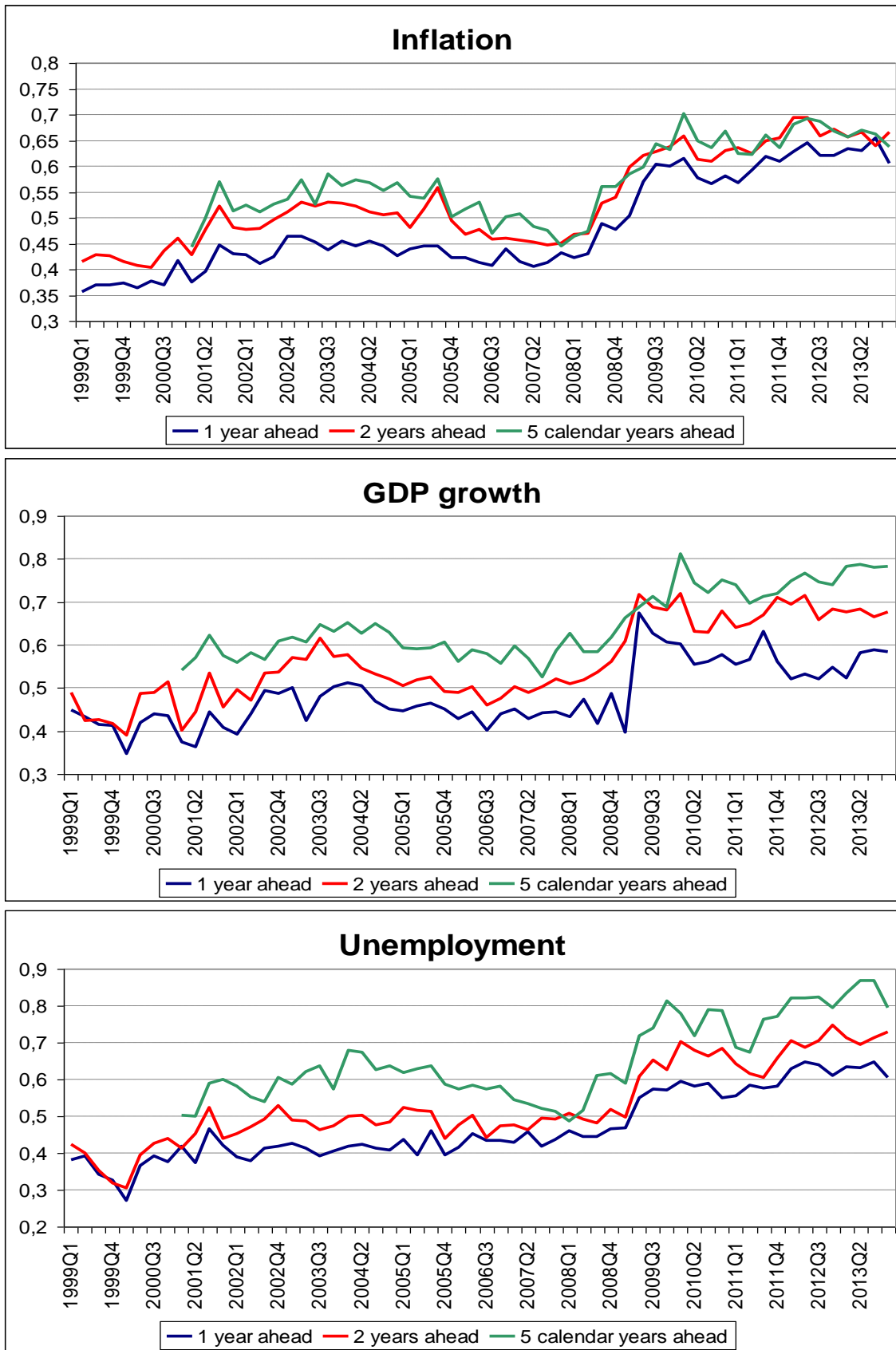
Figure 10: Implied RMSV of the individual density forecasts

Inflation

GDP growth

Unemployment
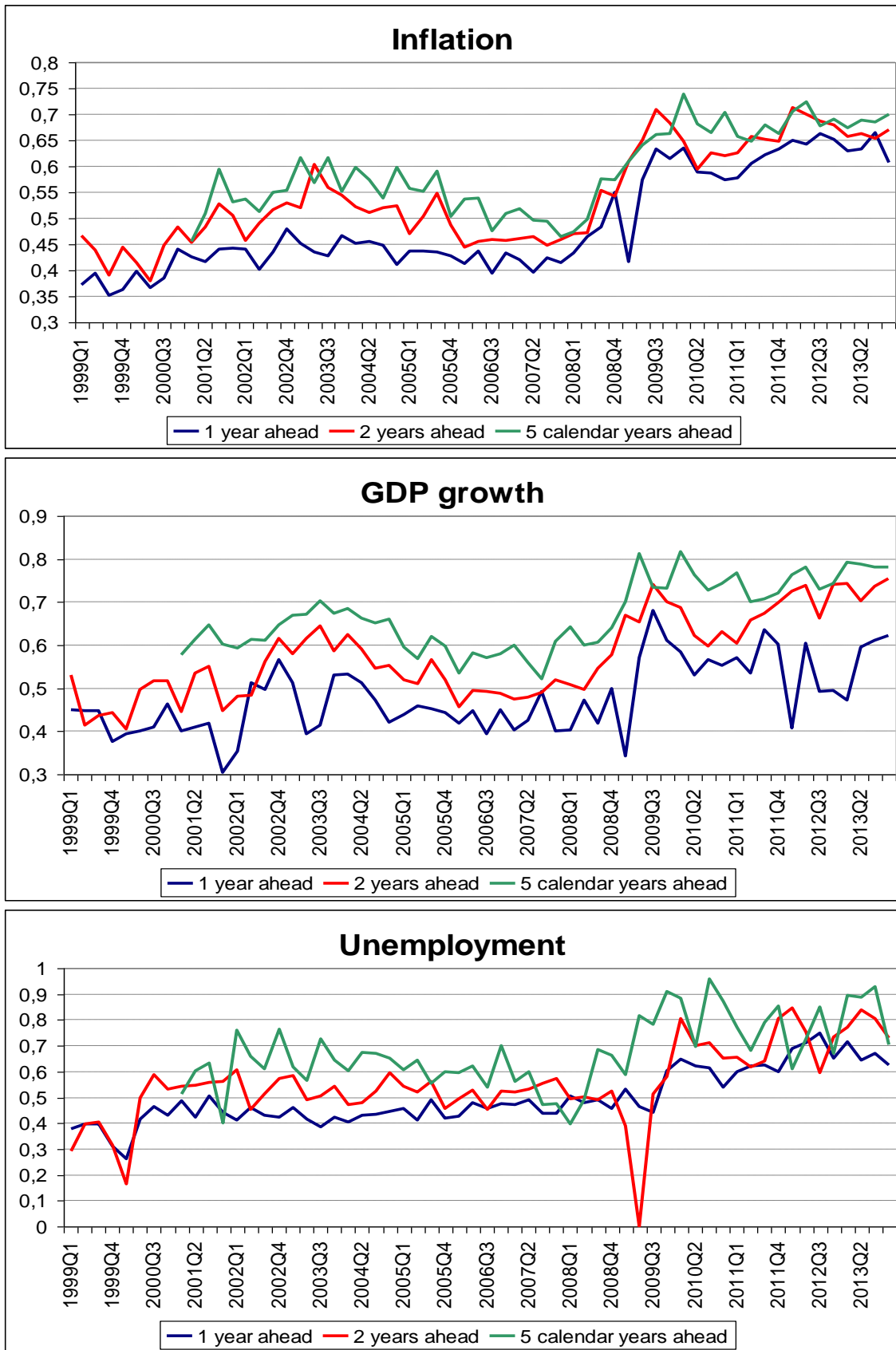
Figure 11: Implied RMSV of the individual density forecasts under different assumptions

GDP growth (one year ahead)

Midpoint — Normal — Beta — Mod midpoint

GDP growth (two years ahead)

Midpoint — Normal — Beta — Mod midpoint

GDP growth (five calendar years ahead)

Midpoint — Normal — Beta — Mod midpoint

Unemployment (one year ahead)

Midpoint — Normal — Beta — Mod midpoint

Unemployment (two years ahead)

Midpoint — Normal — Beta — Mod midpoint

Unemployment (five calendar years ahead)

Midpoint — Normal — Beta — Mod midpoint

Figure 12: Median IQR of the individual density forecasts
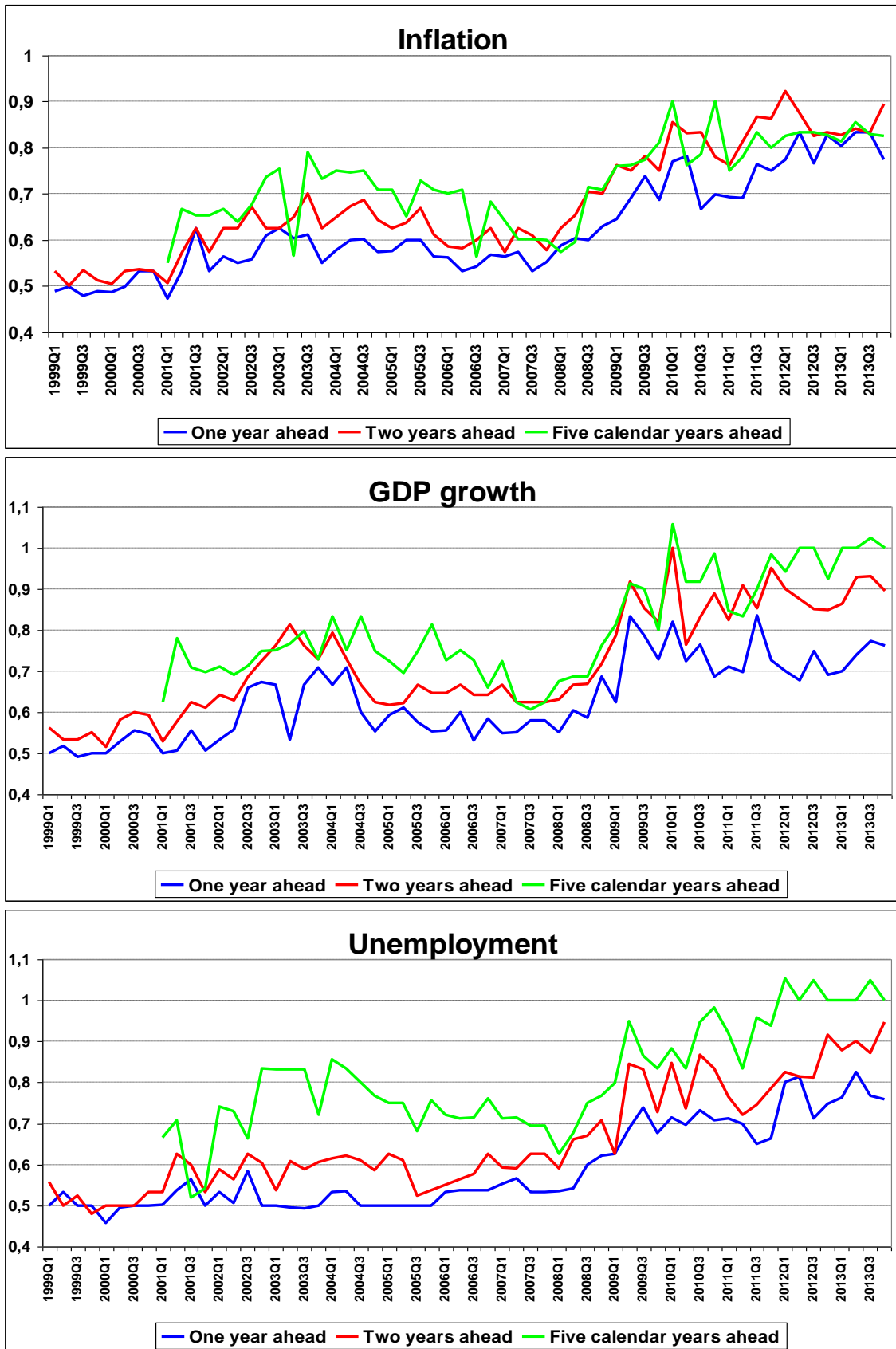
Figure 13: Average entropy of the individual density forecasts

Inflation

GDP growth

Unemployment
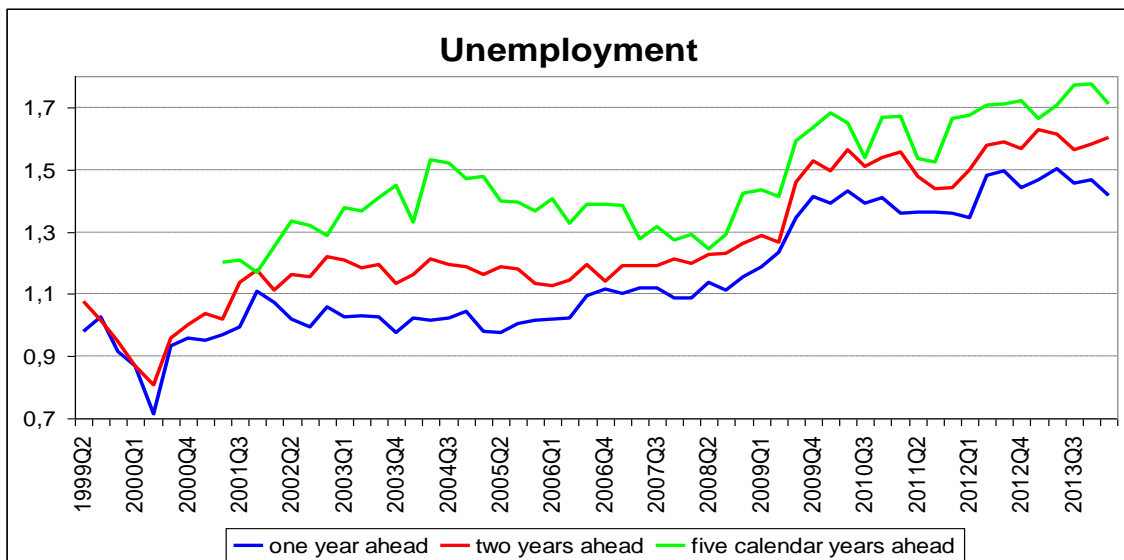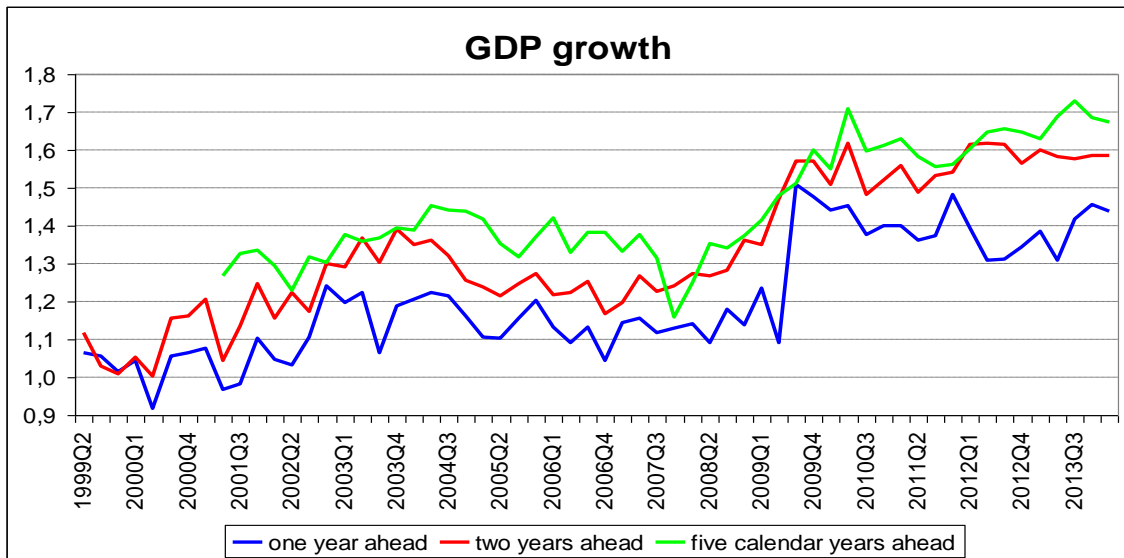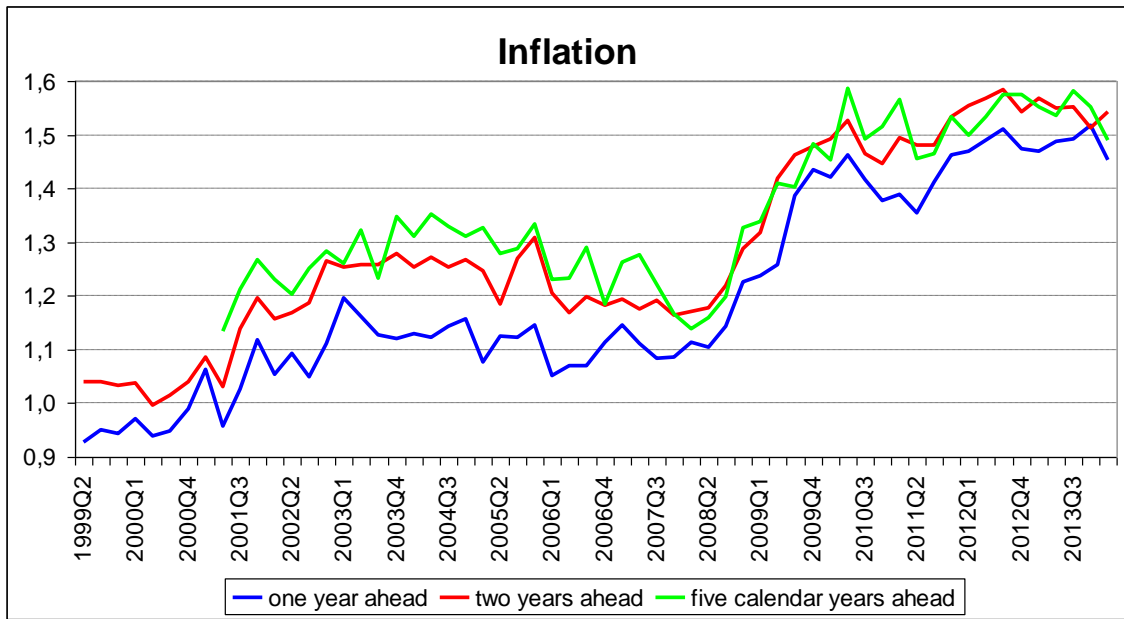
Figure 14: Illustration of the concept of entropy with two intervals
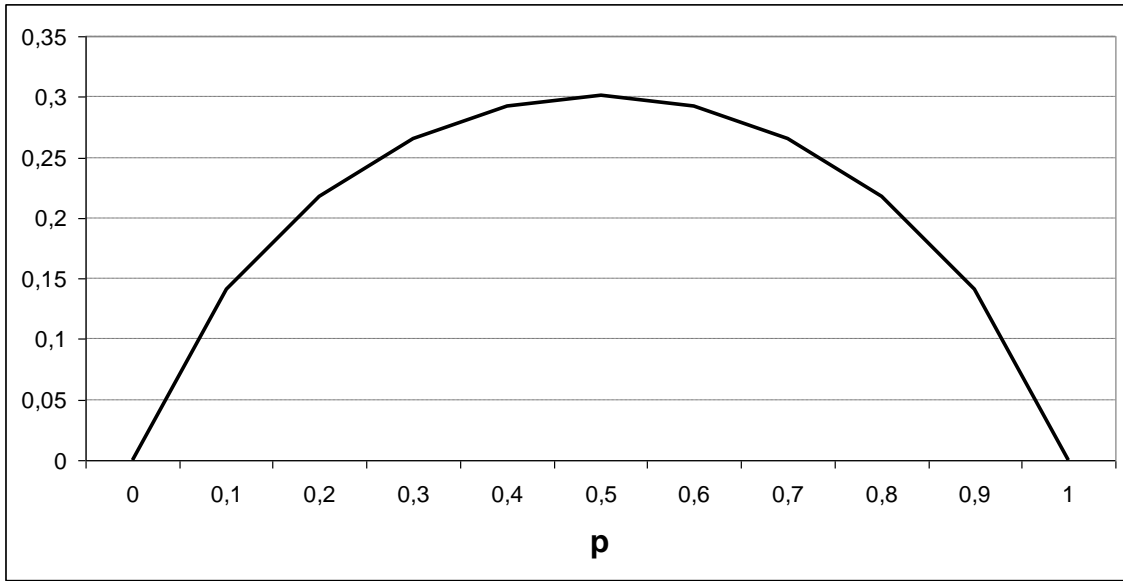
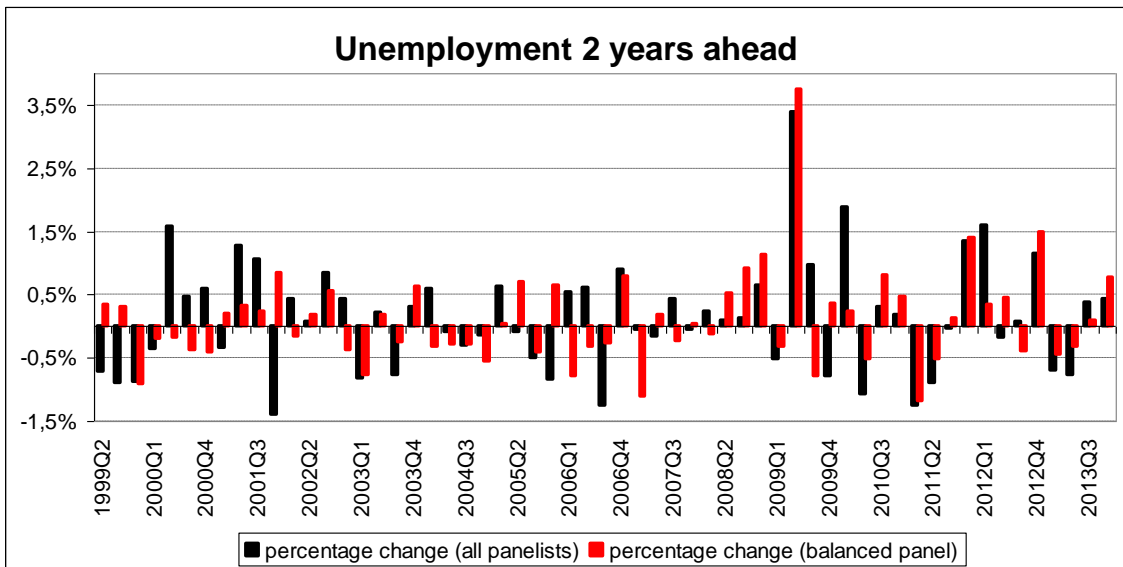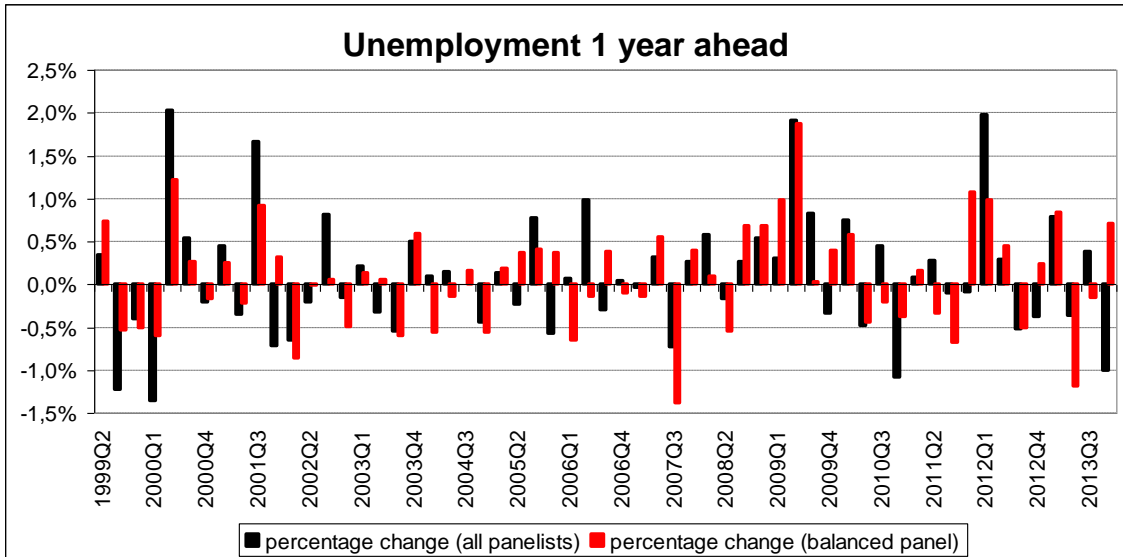**Figure 15: Lorenz curves for selected individual density forecasts**
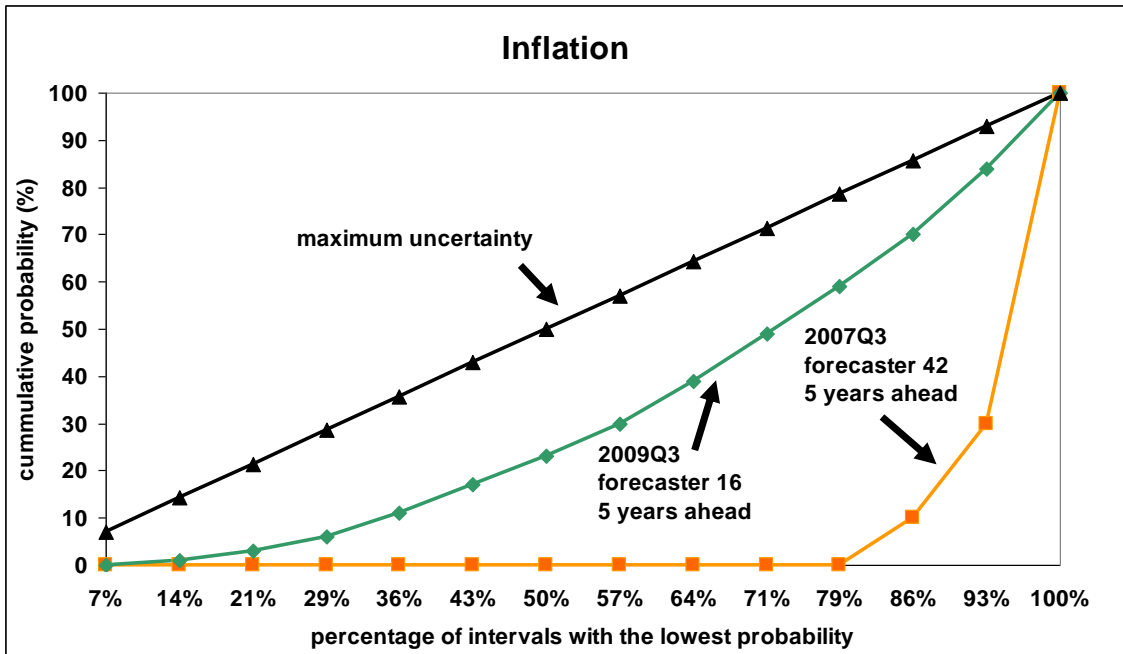
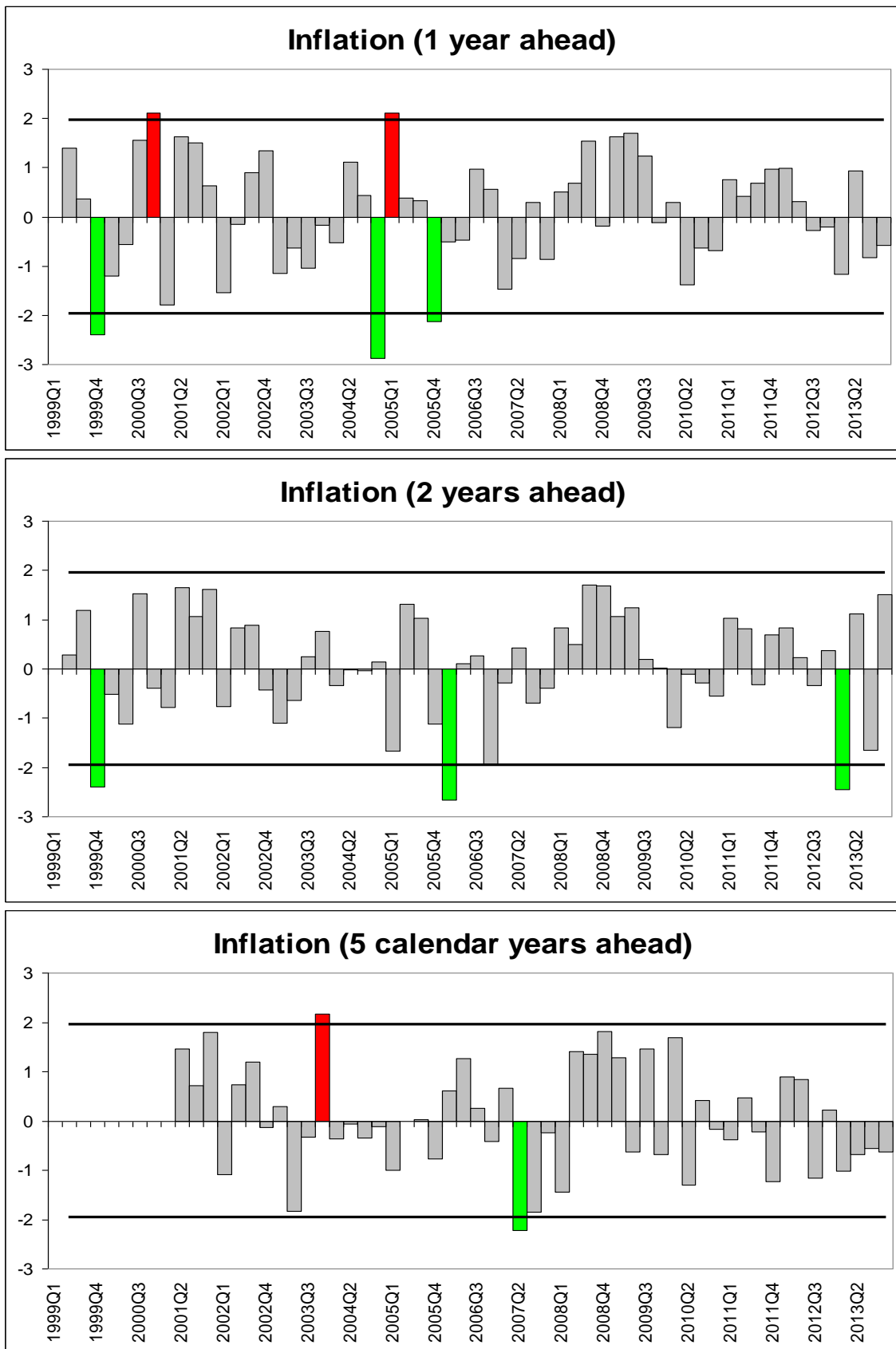**Figure 16: Average Gini indices of individual density forecasts**
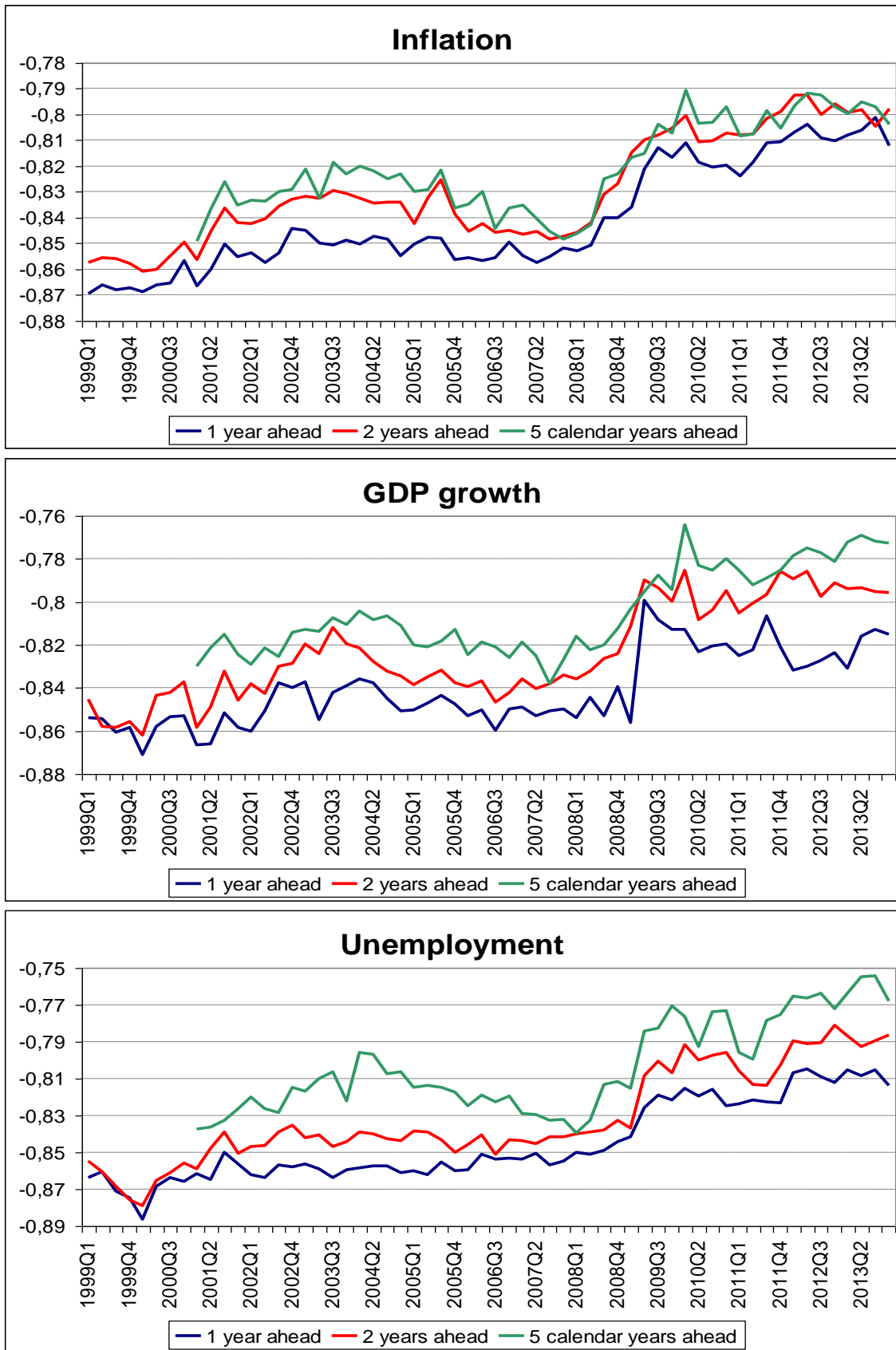
Figure 17: Percentage changes in the average Gini index of individual density forecasts
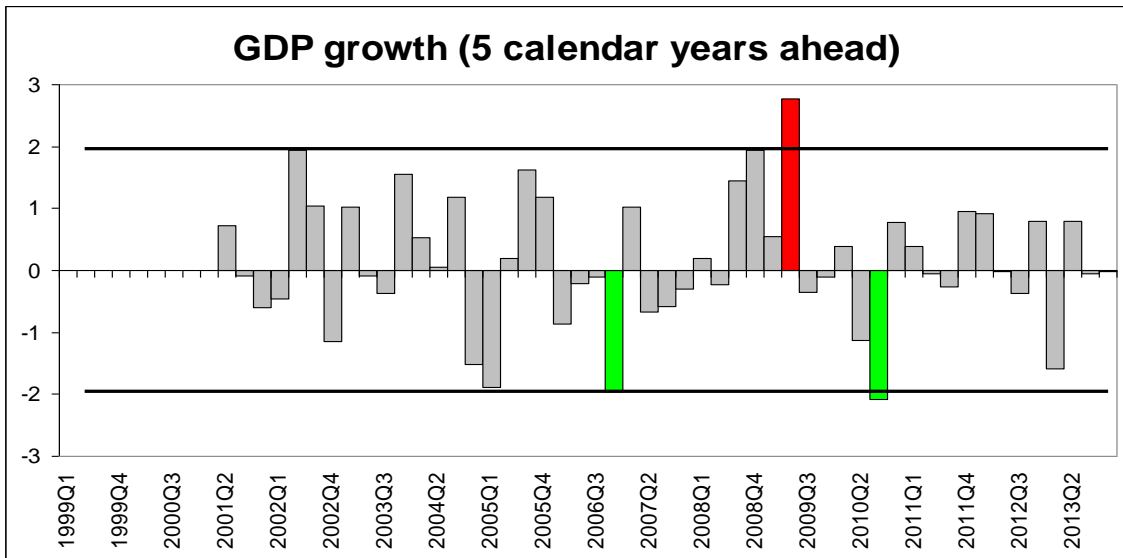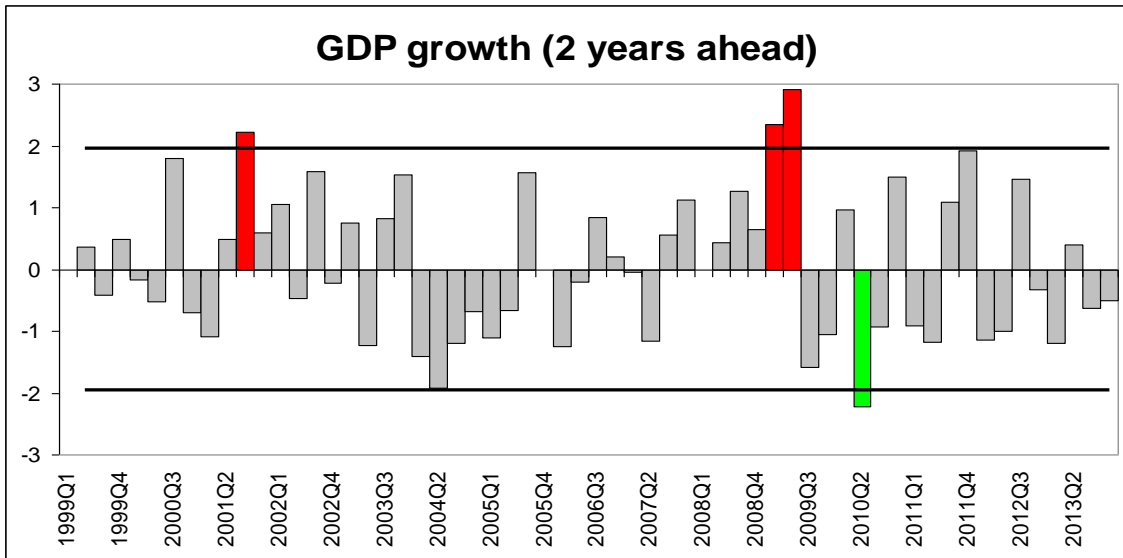
GDP growth 1 year ahead

■ percentage change (all panelists) ■ percentage change (balanced panel)

GDP growth 2 years ahead

■ percentage change (all panelists) ■ percentage change (balanced panel)

GDP growth 5 calendar years ahead

■ percentage change (all panelists) ■ percentage change (balanced panel)

Unemployment 1 year ahead

■ percentage change (all panelists)　■ percentage change (balanced panel)

Unemployment 2 years ahead

■ percentage change (all panelists)　■ percentage change (balanced panel)

Unemployment 5 calendar years ahead

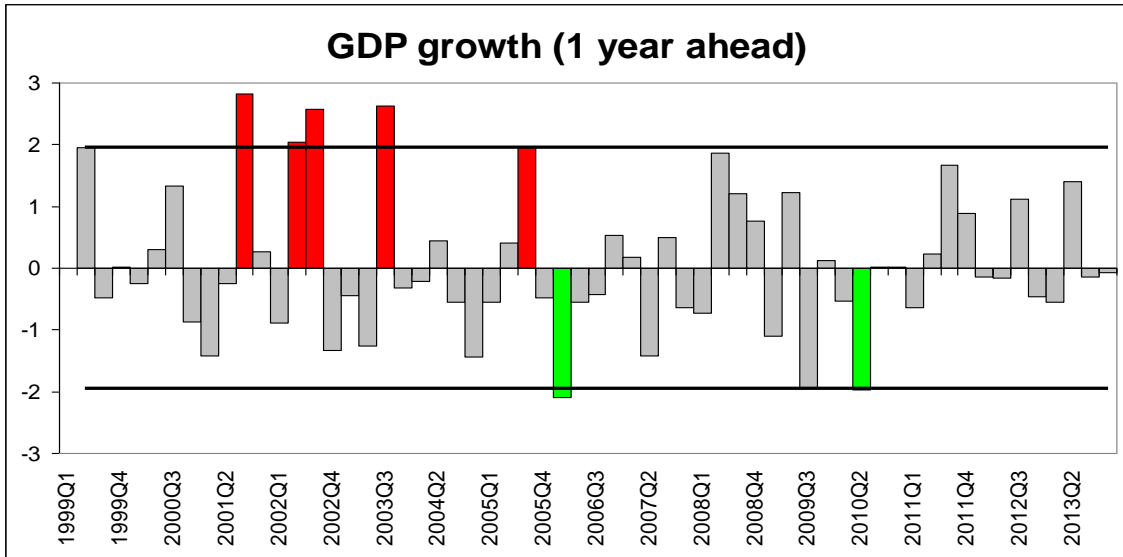■ percentage change (all panelists)　■ percentage change (balanced panel)
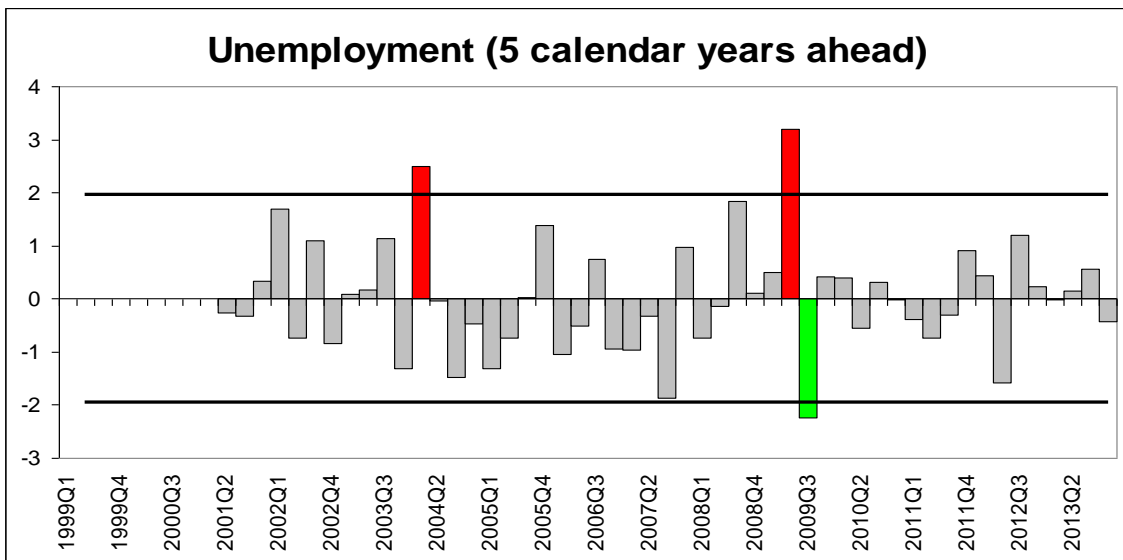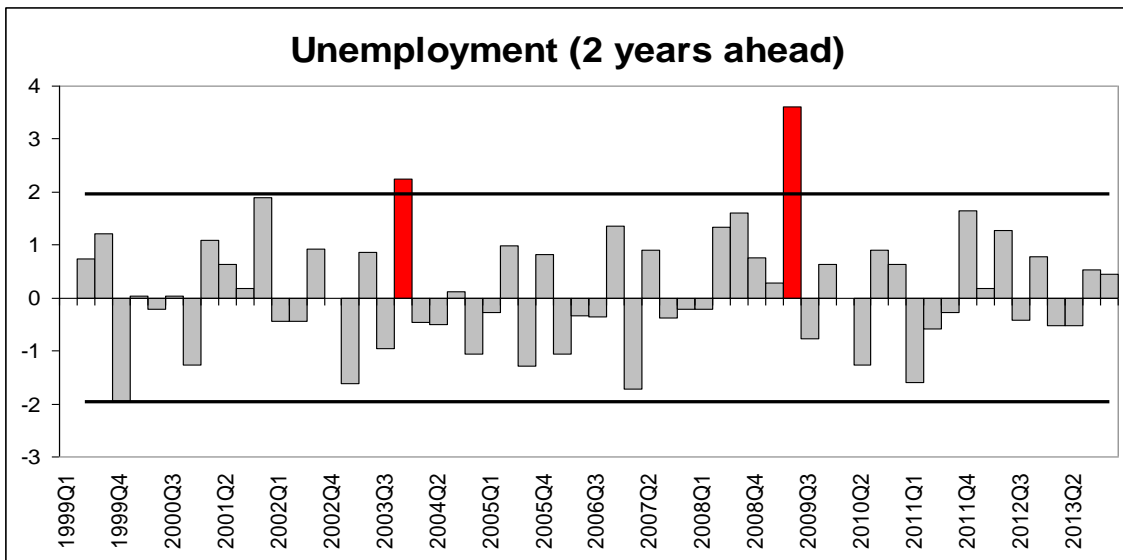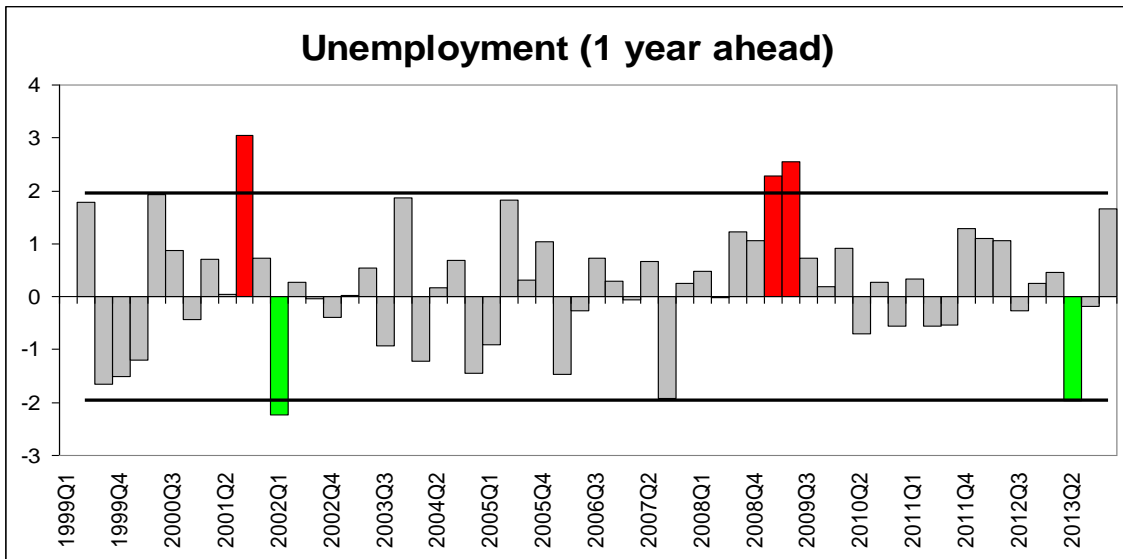
Figure 18: Results of the Wilcoxon signed-rank test of changes in the average Gini index of the individual density forecasts from one quarter to the next

GDP growth (1 year ahead)

GDP growth (2 years ahead)

GDP growth (5 calendar years ahead)

Unemployment (1 year ahead)

Unemployment (2 years ahead)

Unemployment (5 calendar years ahead)

Figure 19: Combined results of the Wilcoxon signed-rank test of changes in the average Gini indices of the individual density forecasts from one quarter to the next

All variables, all horizons