

University of Heidelberg

Department of Economics



Discussion Paper Series | No. 571

**A Multivariate Analysis of Forecast Disagreement:
Confronting Models of Disagreement with SPF Data**

Jonas Dovern

September 2014

A Multivariate Analysis of Forecast Disagreement: Confronting Models of Disagreement with SPF Data^{*}

Jonas Dovern[†]

Alfred-Weber-Institute for Economics, Heidelberg University

September 5, 2014

Abstract

This paper documents multivariate forecast disagreement among professional forecasters of the Euro area economy and discusses implications for models of heterogeneous expectation formation. Disagreement varies over time and is strongly counter-cyclical. Disagreement is positively correlated with general (economic) uncertainty. Aggregate supply shocks drive disagreement about the long-run state of the economy while aggregate demand shocks have an impact on the level of disagreement about the short-run outlook for the economy. Forecasters disagree about the structure of the economy and the degree to which individual forecasters disagree with the average forecast tends to persist over time. This suggests that models of heterogeneous expectation formation, which are currently not able to generate those last two features, need to be modified. Introducing learning mechanisms and heterogeneous signal-to-noise ratios could reconcile the benchmark model for disagreement with the observed facts.

JEL Classification: E27, E37, E52

Keywords: Macroeconomic expectations, forecasts, noisy information, survey data, disagreement

^{*}I am grateful for valuable discussions about the paper with Zeno Enders, Thomas Eife, Ulrich Fritsche, Jochen Güntner, Matthias Hartmann, Nils Jannsen, Pia Pinger, and Jörg Rieger. The paper has also greatly benefitted from comments made after presentations at Heidelberg University and at the 29th annual congress of the European Economic Association in Toulouse.

[†]Correspondence to: Jonas Dovern, Alfred-Weber-Institute for Economics, Heidelberg University, Bergheimer Str. 58, 69115 Heidelberg, Germany. Phone: +49-6221-54-2958. E-mail: jonas.dovern@awi.uni-heidelberg.de.

1 Introduction

Given the public availability of macroeconomic data, it is surprising to observe high degrees of disagreement across agents when it comes to forecasting the future state of the economy. The degree of forecast disagreement for several macro variables and different countries has been documented by a number of recent papers including [Mankiw et al. \(2003\)](#), [Carroll \(2003\)](#), [Lahiri and Sheng \(2008\)](#), [Banerghansa and McCracken \(2009\)](#), [Rich and Tracy \(2010\)](#), [Patton and Timmermann \(2010\)](#), [Coibion and Gorodnichenko \(2012\)](#), [Dovern et al. \(2012\)](#), [Andrade and Le Bihan \(2013\)](#) and [Andrade et al. \(2013\)](#). However, there are two aspects of forecast disagreement that have largely been neglected by the existing literature. First, all papers—with the exception of the study by [Banerghansa and McCracken \(2009\)](#)—ignore the multivariate dimension of macroeconomic forecast disagreement.^{1,2} Second, none of the existing papers looks at the persistence of disagreement of individual forecasters with the prevailing average predicted state of the economy.

In this paper, I address these two points by analyzing forecast disagreement among professional business-cycle analysts in the Euro area from a multivariate perspective. My analysis is relevant for two main reasons. First, a better understanding of the degree to which agents in the economy disagree about the future—and about factors that influence disagreement—is of direct interest for economic policy. E.g., one goal of all major central banks is to anchor inflation expectations at some specific target value and, thus, ideally they would like to see disagreement about future inflation rates to disappear. Second, knowledge about the dynamics of overall disagreement and the persistence of the relative level of individual disagreement helps researchers to better understand how macroeconomic expectations are formed and how they should be modelled.

A suitable multivariate approach for measuring disagreement can capture aspects of disagreement that the conventional (univariate) analyses have neglected. As [Banerghansa and McCracken \(2009, p.2\)](#) write, “it is reasonable to assume that [forecasters]

¹[Banerghansa and McCracken \(2009\)](#) analyze the level of multivariate disagreement among members of the Federal Open Market Committee (FOMC) of the Federal Reserve Bank and relate the individual level of disagreement to several characteristics of the members (such as their voting status). Thus, they look at forecasts made by a very particular type of agents. In addition, the time dimension of their data set is rather limited so that it does not allow one to analyze, for instance, business-cycle effects on disagreement.

²In fact, [Andrade et al. \(2013\)](#) build a multivariate model of expectation formation. They do not, however, use any information about multivariate disagreement to calibrate and evaluate their model.

construct their vectors of forecasts in a congruent fashion that jointly describes their view of the economy”. Consequently, when measuring disagreement among forecasters, they argue, a multivariate perspective is more appropriate than treating disagreement about the future path of each scalar variable separately. By using an appropriate multivariate approach it is also possible to capture disagreement about correlations between different variables—for instance about the nature of the Phillips curve relationship between inflation and the unemployment rate. Clearly, empirical evidence on this aspect would be informative about the appropriateness of the assumption that all agents agree on the structure of the economy, which is often made in models of disagreement (see, e.g., [Andrade et al., 2013](#)).

To see why it makes sense to take a multivariate perspective for analyzing forecast disagreement in the context of vector-valued forecasts, a look at the following simple bivariate example is enlightening. Consider the set of forecasts made for two random variables X and Y in Figure 1. The black dots are generated from a bivariate normal distribution with a high positive correlation. This positive correlation reflects the fact that forecasters, on average, seem to think that higher values of X go along with higher values of Y . Thus, they tend to agree on the model but disagree on the particular scenario that will materialize. Now consider the four forecasts indicated by the red squares. Based on univariate measures of disagreement or multivariate ones that do not take into account correlations between variables (such as the Euclidean distance) all of these forecasts imply the same level of disagreement. Given the entire set of forecasts, however, it is reasonable to argue that an appropriate multivariate measure of disagreement should indicate a higher level of disagreement for the forecasts in the lower right and upper left square relative to the level of disagreement associated with the lower left and the upper right forecast.

[Figure 1 about here.]

Speaking in economic terms, a large general dispersion of forecasts (neglecting correlations) indicates that forecasters disagree strongly about the most likely scenario for the future state of the economy. While some forecasters could for instance expect a strong expansion, others could expect more moderate growth or even an outright recession. Such divergent views about the business-cycle prospects are likely to lead to large forecast disagreement for most macroeconomic variables. As has been seen, however, another aspect of disagreement in the multivariate context is that of the correlation between forecasts. This property of forecasts reflects the degree to which forecasters

agree or disagree on the model that describes how the economy functions. If forecasts for two particular variables show a high degree of correlation, forecasters use similar models to generate their forecasts.³ On the contrary, if those forecasts are uncorrelated, the forecasters' views on the proper model for describing the behavior of the two variables differ greatly, i.e., they disagree about the model of the economy. Both aspects are relevant for determining the overall level of disagreement.

Different models have been proposed to model forecast disagreement among agents, i.e., to generate forecast disagreement within one particular model with heterogeneous expectation formation. In general, these models involve some form of imperfect information that has idiosyncratic elements and/or agent-specific beliefs to generate heterogeneity across agents. One type of such a model is the noisy-information model of [Sims \(2003\)](#) and [Woodford \(2003\)](#). In this model, agents receive either noisy signals or are not able to fully process all the information they receive. Agents then use Bayes' rule to optimally update their forecasts. Due to the fact that they receive different signals (or interpret signals differently), their forecasts can diverge and disagreement among agents occurs. [Andrade et al. \(2013\)](#) generalizes this setup to allow the imperfectly observed state of the world to be the sum of a transitory and a permanent component; thereby they are able to match the empirical fact that forecasters disagree also about the very long-run outlook. A different type of model is proposed by [Lahiri and Sheng \(2008\)](#) and [Patton and Timmermann \(2010\)](#) who use heterogeneous prior beliefs to replicate the fact that agents also disagree about the long-run outlook.⁴ All of these existing models share the assumption that agents agree on the structure of the model/economy. Furthermore, all of the basic models assume that agents are alike⁵ and they are, thus, not able to generate persistent differences in the characteristics of individual forecasts. As I will show in this paper, both of these features are rejected empirically.

³To be more precise, they use models that imply similar linear relationships between the variables. Theoretically, these could be generated by two or more different (potentially non-linear) structural models. In addition, it should be stressed that this interpretation rules out the possibility of anticipated shocks. (If, for instance, a subgroup of forecasters anticipated an aggregate supply shock when formulating their multivariate forecasts, these forecasts would imply a different correlation between growth and inflation than the forecasts made in non-anticipation of such a shock.) I will rule out the possibility of such anticipated shocks throughout this paper.

⁴A third type of models that has been proposed to explain disagreement among agents is the sticky-information model of [Mankiw and Reis \(2002\)](#). It has been empirically rejected, however, as an appropriate model to describe the behavior of professional macro forecasters by [Dovern \(2013\)](#), [Andrade and Le Bihan \(2013\)](#), and [Dovern et al. \(2014\)](#).

⁵That means their idiosyncratic signals and/or priors are drawn from the same distributions without any persistence over time.

The main findings of the paper can be summarized as follows. First, I show that multivariate disagreement is strongly time-varying and that disagreement about the long-run is usually higher than disagreement about the near future—features that are consistent with the extended noisy information model of [Andrade et al. \(2013\)](#) or the models with individual priors about the long-run ([Lahiri and Sheng, 2008](#); [Patton and Timmermann, 2010](#)). Second, I show that disagreement about the correlation between different variables is very high in general, suggesting that forecasters do not agree on how the economy works. At the same time, movements in those components do not explain a large fraction of variation of overall disagreement at the business-cycle frequency, suggesting that the degree to which forecasters disagree about macroeconomic mechanisms does not vary a lot over different phases of the business cycle. Third, I show that the relative level of disagreement of individual forecasters displays some persistence over time.⁶ Fourth, I show that overall disagreement is increasing in the level of prevailing macroeconomic uncertainty—a feature that is consistent with existing models of disagreement (if uncertainty is measured by the signal-to-noise ratio). Finally, I conclude that the assumption of common knowledge about the true structure of the model/economy should be abandoned in theoretical models of disagreement⁷. Furthermore, future models should allow for persistence in the ranking of relative levels of individual disagreement.

The remainder of this paper is organized as follows. Section 2 explains the measures that I use to estimate the degree of multivariate disagreement. Section 3 discusses the data set that I use. Section 4 documents the evolution of disagreement over time and analyzes i) the persistence of levels of individual disagreement, ii) the comovement of disagreement with different proxies for (macro-)economic uncertainty, and iii) the reaction of disagreement to different macroeconomic shocks. Section 5 elaborates on the implications that my empirical results have for the future development of theoretical models of forecast disagreement. Section 6 concludes.

2 Measuring Multivariate Disagreement

In this paper, I focus on two different (though related) multivariate measures of forecast disagreement. The first measures the disagreement of individual forecasters relative to

⁶In other words, some forecasters tend to persistently publish forecasts that are not in line with the current average forecast while others tend to persistently publish forecasts which are very similar to the average forecast. The concept will be formally defined below.

⁷[Andrade et al. \(2013\)](#) also suggest this feature as a potential future generalization of their model.

the prevailing average level of disagreement in each period, i.e., this measure indicates whether a particular forecaster in a particular time period reports a forecast that is rather similar to the average forecast or very different from what other forecasters are predicting. The second measure reflects the overall level of disagreement among the group of forecasters in a particular time period; it measures how dispersed the set of observed multivariate forecasts is at each point in time and can be used to compare the prevailing level of disagreement over time.

Banternghansa and McCracken (2009) suggest to use the Mahalanobis distance to measure relative disagreement for each individual forecasters and to use the determinant of the cross-sectional sample covariance matrix of the individual vector forecasts as a measure of overall multivariate disagreement.⁸ Let $y_{i,t+h|t} = [y_{i,t+h|t}^1, y_{i,t+h|t}^2, \dots, y_{i,t+h|t}^k]'$ denote the forecast vector published by agent i at time t with a forecast horizon of h periods and denote by $\bar{y}_{t+h|t}$ the average of these forecasts. Given the sample covariance of the individual forecasts $S_{t+h|t} = N_{t+h|t}^{-1} \sum_{i=1}^{N_{t+h|t}} (y_{i,t+h|t} - \bar{y}_{t+h|t})(y_{i,t+h|t} - \bar{y}_{t+h|t})'$ with $N_{t+h|t}$ being the number of observed h -step-ahead forecasts at time t , the Mahalanobis distance for a particular forecaster is given by

$$d_{i,t+h|t} = \sqrt{(y_{i,t+h|t} - \bar{y}_{t+h|t})' S_{t+h|t}^{-1} (y_{i,t+h|t} - \bar{y}_{t+h|t})}. \quad (1)$$

One property of this measure is that it rescales the absolute level of disagreement separately for each time period and forecast horizon (through the normalization by $S_{t+h|t}^{-1}$). Thereby, it provides a measurement of relative disagreement of a particular forecaster (compared to the average level of disagreement that is prevailing in a particular time period). It allows to rank forecasters according to their corresponding degrees of disagreement with the prevailing average forecast.

A measure of the overall level of disagreement prevailing at time t with respect to forecasts with a horizon of h is given by the square root of the determinant of $S_{t+h|t}$:

$$D_{t+h|t} = \sqrt{\det(S_{t+h|t})} \quad (2)$$

This measure is increasing in the cross-sectional variances for each particular scalar forecast, i.e., the elements on the main diagonal of $S_{t+h|t}$ and decreasing in the absolute value of each of the cross-sectional correlations between any pair of two scalar

⁸In some text passages, I use the term *overall disagreement* as a synonym for *multivariate disagreement*.

forecasts, i.e., the off-diagonal elements of $S_{t+h|t}$.⁹ It measures the absolute level of multivariate disagreement and, thus, is useful for comparisons across time and across different forecast horizons.

3 Data

I use data on forecasts made by professional business-cycle analysts from the Survey of Professional Forecasters (SPF) that has been published at a quarterly frequency by the European Central Bank (ECB) since the first quarter of 1999. The data set covers forecasts from a wide range of professional forecasters—mostly from research institutes and financial institutions. The panel is so homogeneous, though, that it is unlikely that major differences in incentive or market structures trivially explain much of the disagreement across forecasters.¹⁰ The data set contains forecasts for the growth rate of real gross domestic product (g_t), the inflation rate (π_t), and the unemployment rate (u_t), i.e., throughout the paper $k = 3$ and $y_{i,t+h|t} = [g_{i,t+h|t}, \pi_{i,t+h|t}, u_{i,t+h|t}]'$. The SPF data include information about forecasts of different types (fixed-event as well as fixed-horizon forecasts) and with various forecast horizons. In this paper, I concentrate on the fixed-horizon forecasts with the shortest forecast horizon available (1-year-ahead forecasts, $h = 4$) and those with the longest forecast horizon available (5-years-ahead forecasts, $h = 20$).¹¹

The sample covers the forecast periods from 1999q1 to 2014q2. Given the focus of this paper, I consider only such observations for which forecasts for all three variables are provided by a particular forecaster. The panel is unbalanced because some forecasters left the survey at some point in time and others entered after the start of the SPF in 1999; in addition, forecasters are not required to respond for each wave of the survey. The average number of long-term forecasts observed at each point in time is 34.7 and

⁹A theoretical drawback of this measure—which, however, is unlikely to become relevant in practice—is the fact that it approaches its minimum 0 if only one of the elements on the main-diagonal of $S_{t+h|t}$ approaches zero and also if only one of the correlations implied by the off-diagonal elements approaches 1.

¹⁰A list of participating institutions can be found on the ECB website (<http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html>).

¹¹Strictly speaking, the 5-years-ahead forecasts are fixed-event forecasts made for the annual average of the forecast variables in a particular target year. This target year is changing in such a way that the forecast horizon varies between 21 and 18 quarters. Given the very long forecast horizon it is reasonable to expect that forecasters use their estimates for the unconditional mean of each variable as their forecasts; in this case the forecasts are likely to be unaffected by changes of the target year or small variations of the forecast horizon.

the average number of short-term forecasts is 40.8 (Table 1). The volatility of the number of observations across time is rather moderate, but the number of available forecasts tends to decline towards the end of the sample.¹²

[Table 1 about here.]

4 Multivariate Disagreement about the Euro Area Outlook

4.1 Disagreement over Time

In this section, I document the evolution of multivariate disagreement about the future state of the economy in the Euro area among participants of the SPF over time. This assessment is based on the measure defined by equation 2. Figure 2 shows the time series of multivariate disagreement for both forecast horizons. Two main features are visible from the graph. First, disagreement seems to intensify during recessions (2001/02; 2008/09; 2011/12). Second, the level of disagreement about the state of the economy in the very long-run (black line) is usually higher than that about economic conditions in the nearer future (red line)—except during some of the recessionary periods. Interestingly, the higher level of disagreement about the state of the economy in the long-run is caused by higher disagreement about the long-run unemployment rate and higher disagreement about the correlations between the three variables in the long-run.¹³ Disagreement about the long-run inflation rate and the long-run growth rate is, on average, lower than that about the nearer term movements of those variables.

[Figure 2 about here.]

What elements of $S_{t+h|t}$ are driving most of the fluctuations in multivariate disagreements over time? Is it variations in the level of disagreement about future business-cycle scenarios (elements on the main diagonal)? Or is it time-variation in the level of disagreement about the appropriate model of the economy (correlations on the off-diagonal elements)? Figure 3 shows the time series of each of the elements in $S_{t+h|t}$ for both forecast horizons. It is evident that medium-to-low-frequency movements in multivariate disagreement are driven primarily by movements in the cross-sectional standard

¹²In 1999 the survey started with an average number of respondents (reporting short-term forecasts) of 46.5; in 2013 on average only 33.5 forecasters submitted forecasts for all three variables.

¹³The higher disagreement about the correlations between variables is reflected in lower cross-sectional correlations between pairs of long-run forecast relative to correlations based on short-run forecasts.

deviation for each of the three variables while fluctuations of the implied bivariate correlations are of higher frequency. The impression is confirmed by a regression of $D_{t+h|t}$ on the three standard deviations in one case and on the three correlations in the other case. For the short-term forecasts the univariate standard deviations alone explain 85% of the variation of $D_{t+4|t}$ while the correlations alone explain only 8% of that variation. For the long-term forecasts the univariate standard deviations alone explain even 95% of the variation of $D_{t+20|t}$ while the share of variation explained by the correlations alone is up to 22%. Thus, in general variation in the level of disagreement about the right model does not seem to be very important for movements in overall disagreement among forecasters.

[Figure 3 about here.]

But what about the unconditional level of (dis-)agreement about the relationship between the three macroeconomic variables? In general disagreement is quite high, i.e., the observed cross-sectional correlations are relatively small (in absolute values). Table 2 shows the full-sample averages. All correlations have the expected sign. The cross-section of forecasts on average reflect a Phillips curve relationship in the sense that higher growth forecasts tend to go along with higher forecasts for inflation. The correlation is stronger for the short-term forecasts (.16) than for the long-term forecasts (.10). This comes also to no surprise since long-term growth forecasts are likely to reflect assessments of the potential growth rate of the economy which in itself should not affect inflation in the long-run. The implied Phillips curve relationship in terms of an inflation-unemployment trade-off is less pronounced but has the expected sign. The belief of forecasters in the Phillips curve has been documented also for instance by Fendel et al. (2011). Forecasters seem to agree most on the relationship between growth and the unemployment rate (Okun’s law). The average observed correlation based on short-term forecasts is -0.25—and even for long-term forecasts it is -0.17. This finding is in line with evidence reported by for instance Frenkel et al. (2011) or Ball et al. (2014).

[Table 2 about here.]

How did overall disagreement evolve following the Great Recession? Anecdotal evidence suggest that no consensus about the causes and implications of the sharp decline in economic activity in the Euro area has yet emerged. To formally check whether there have been significant shifts in multivariate disagreement after the beginning of

the Great Recession and to analyze which of the components have been driving these shifts, I regress $D_{t+h|t}$ or its components respectively on a constant and two dummy variables. The first dummy variable (*Crisis*) captures shifts that are limited to the period between 2008q3 and 2010q2. The second dummy variable (*Post2008q3*) captures permanent (or at least very persistent) shifts that have occurred after the beginning of the Great Recession.

Signs of an increase in disagreement among the participants of the SPF after 2008 could be twofold. First, forecasters could potentially disagree more on the future scenario that is most likely for the economy while in general agreeing or disagreeing to the same extent about the functioning/structure of the economy. Second, forecasters could potentially disagree more about how the economy works—how growth, inflation and the unemployment rate move together—than before the crisis.

First, the results show that multivariate disagreement about the near future increased substantially and significantly during the Great Recession but decreased back to moderate levels after 2010, which are not significantly higher than those before the crisis (Table 3). Second, multivariate disagreement about the long-term outlook for the Euro area has nearly doubled after 2008 compared to the period before the crisis. Third, the results show that changes in disagreement about the co-movement of variables are not the main drivers of overall disagreement. Only three of the relevant twelve coefficients are significantly different from zero: Disagreement about the co-movement of inflation and the unemployment rate in the short-run temporarily decreased during the Great Recession as did disagreement about the co-movement of growth and inflation in the long-run. Surprisingly, disagreement about the co-movement of growth and the unemployment rate in the long-run decreased during the Great Recession. Comparing the entire post-2008 period to the period before the Great Recessions reveals no significant changes in the level of disagreement about the relationship between the variables at all.

[Table 3 about here.]

Finally, while disagreement about future growth rates temporarily increased during the Great Recession, most of the permanent increase in overall disagreement (both for the short-term and the long-term outlook) is due to an increased dispersion of forecasts for inflation and the unemployment rate. The permanent increase is about 1/3 for both variables and for both forecast horizons. Overall, the observed increase in disagreement seems to be mainly due to an increased uncertainty about the business-cycle scenarios

which are going to materialize rather than due to an increased divergence of opinions about the functioning of the economy.

4.2 Persistence of Levels of Individual Disagreement

Turning to the persistence of the level of relative individual disagreement, it is of interest to re-state what has been found with respect to the persistence of rankings of other moments of the forecast distribution. A high degree of persistence in the ranking of forecasters has been documented with respect to the level of point forecasts (Batchelor, 2007; Patton and Timmermann, 2010; Boero et al., 2014) and with respect to the level of subjectively perceived forecast uncertainty (Bruine De Bruin et al., 2011; Boero et al., 2014). Both observations are in contrast to the prediction of all models that have been proposed to model disagreement and that were discussed in the introduction to this paper. To improve the foundation for these models it is also of interest to document the degree of persistence in the ranking of forecasters with respect to the individual level of disagreement. The currently used models of heterogeneous expectation formation imply a low degree of such persistence.

As a first inspection of the data consider Figure 4. The blue dots refer to observations corresponding to those 10 forecasters that have the highest average rank in terms of disagreement over the entire sample.¹⁴ The red dots refer to observations corresponding to those 10 forecasters that have the lowest average rank in terms of disagreement over the entire sample. Both plots (for long- and short-term forecasts) suggest that there is substantial persistence in the forecasters' relative level of disagreement. The disagreement rankings look less persistent, however, than most rankings analyzed in the papers cited in the previous paragraph. The plots also indicate an asymmetry in the sense that those forecasters that disagree only very little on average do so very persistently whereas the relative level of disagreement of those forecasters that disagree a lot on average is more volatile.

[Figure 4 about here.]

A measure to formally analyze the persistence of rankings over time is the so-called Kendall coefficient of concordance (Kendall, 1948). It measures the ratio of the sum of squared mean deviations of the observed average ranks (computed across time) to

¹⁴These “extreme” forecasters are selected among the subset of forecasts for which at least 12 observations are available.

the maximum possible value. The statistic, $0 \leq W \leq 1$, is equal to 0 if the rankings change randomly over time—and consequently all forecasters have the same average rank—and it is equal to 1 if the ranking is perfectly stable over time.¹⁵

Table 4 shows the Kendall coefficient of concordance of the rankings of the individual levels of disagreement, $d_{i,t+h|t}$, for the two forecast horizons and different samples.¹⁶ The first line corresponds to the full-sample statistic while the lower lines correspond to the pre-crisis and post-crisis periods respectively. The persistence of this ranking is generally higher in case of the long-term forecasts (.2 based on the full sample) than in case of the short-term forecasts (.14). Thus, there is a stronger tendency in case of long-term forecasts that some forecasters tend to deviate always a lot from the consensus outlook on the future state of the economy while others tend to closely agree to this average view most of the time. A second observation is that since the beginning of the Great Recession the persistence of the ranking based on disagreement about the short-run outlook has increased substantially relative to the time period before 2008. The measure of concordance increased from .16 to .25.¹⁷

[Table 4 about here.]

In general, the persistence of the relative level of disagreement across forecasters is lower, however, than the persistence of relative levels of point forecasts or the persistence of reported forecast uncertainty as found by Boero et al. (2014) based on a similar data set from Great Britain. They report Kendall coefficients of concordance of roughly .4 (forecast uncertainty) and between .2 and .5 (point forecasts). Based on the data used in this paper I obtain similar results: The persistence of relative forecast uncertainty over time is much higher for both forecast horizons (.3 to .65) and for long-term point forecasts (.27 to .44). For short-term point forecasts (.13 to .19) it is roughly the same as found for the disagreement ranking.

¹⁵Under the null hypothesis of randomly changing rankings over time and the assumption of no missing data the random variable $(T - 1)NW$ (where T denotes the number of time periods and N is the (fix) number of ranked individuals/items) approximately follows a χ^2 distribution with $T - 1$ degrees of freedom.

¹⁶In contrast to the assumption usually made for the computation of Kendall's coefficient of concordance, the panel data used in this study is subject to missing values. I tackle this problem in two steps. First, I only consider those forecasters in the computation of relative rankings with at least 12 reported forecasts. Second, I use the adjusted procedure to compute W which is used by Boero et al. (2014).

¹⁷Note that a smaller value of Kendall's coefficient of concordance for the full-sample compared to the measure for any of the sub-samples is perfectly fine: A ranking could be perfectly stable over two different sub-samples ($W = 1$) while every individual has the same average rank over the full sample ($W = 0$).

4.3 Disagreement and Uncertainty

Disagreement among agents is usually considered to be one component of aggregate forecast uncertainty (see, e.g., [Giordani and Söderlind \(2003\)](#), [Lahiri and Sheng \(2010\)](#) or [Bachmann et al. \(2013\)](#)). However, it is also appealing to think of disagreement as a function of aggregate uncertainty. If, for instance, higher macroeconomic uncertainty lowers the signal-to-noise ratio, forecast disagreement among agents will rise ([Coibion and Gorodnichenko, 2012](#)). [Dovern et al. \(2012\)](#) provide empirical evidence based on survey data from all G7 countries that disagreement among professional forecasters is increasing in the conditional volatility of the forecast variable. So far however, the link between uncertainty and disagreement has been analyzed in the univariate context only.

I use a simple regression framework to analyze the correlation between multivariate disagreement and different measures of general (economic) uncertainty. The model is of the following form

$$D_{t+h|t} = \alpha + \beta D_{t-1+h|t-1} + \gamma' U_t + \varepsilon_{t+h|t}, \quad (3)$$

where the lagged value of multivariate disagreement is included to capture persistence in the absolute level of disagreement among forecasters, U_t contains one or more proxies of general (economic) uncertainty, and $\varepsilon_{t+h|t}$ is an *iid* error term. I use four different measures of uncertainty: The policy uncertainty index provided by [Baker et al. \(2014\)](#), the realized volatility of equity markets (based on the EuroStoxx50), the realized volatility of oil prices, and the realized volatility of the exchange rate of the Euro against the US-Dollar. Initially, I use each of these measures in separate regressions. All uncertainty indicators are pooled into one regression in the fifth model for each forecast horizon. The results are given in [Table 5](#).

The regression results show that disagreement about the future state of the economy is quite persistent. The coefficients corresponding to the lagged values of disagreement are highly significant and also relatively large. They are larger for short-term disagreement (about .5) relative to the case based on long-term forecasts (about .2 to .3). Looking at the correlation between the uncertainty proxies and my measure of multivariate disagreement reveals two interesting results. First, in most of the regressions that include only one measure of uncertainty the relevant coefficients are highly significant indicating a positive correlation between overall disagreement about the future state of the economy and the used measures. Second, the indicator for policy uncer-

tainty is the only proxy that shows no significant correlation with $D_{t+4|t}$ while it is the only proxy that remains significant when all proxies are included in the regression with $D_{t+20|t}$ as the dependent variable. This shows that uncertainty about the direction of (economic) policy is not relevant for the level of disagreement about the short-run outlook for the economy. Furthermore, this shows that policy uncertainty is the main driver of disagreement about the long-run prospects for the economy.¹⁸ Overall, my results suggest that a sizable fraction of variation in multivariate disagreement can be explained by movements in (economic) uncertainty.

[Table 5 about here.]

4.4 Excursion: Reaction to Structural Shocks

In the previous sections I have documented that multivariate disagreement among forecasters is strongly time-varying and correlated for instance with certain measures of economic uncertainty. Even though existing models of disagreement contain no clear notion of different macroeconomic shocks (such as aggregate supply or demand shocks), it is, eventually, interesting to know which structural shocks drive movements in disagreement among forecasters.¹⁹ Given clear results, disagreement among forecasters could, for instance, be used as an observable indicator that might help to identify certain unobservable structural shocks in real-time. To analyze this aspect I resort to a small-scale standard macroeconomic vectorautoregressive model (VAR).

The VAR is used to model a vector of quarterly measured variables, which is given by $x_t = [gdp_t, cpi_t, i_t, D_{t+h|t}]'$, where $D_{t+h|t}$ is given by equation 2 with either $h = 4$ or $h = 20$, gdp_t is the log of real gross domestic product, cpi_t is the log of the (seasonally adjusted) consumer price index, and i_t is main monetary policy interest rate set by the ECB. The reduced form VAR is given by

$$A(L)x_t = u_t, \tag{4}$$

where $A(L) = I - A_1L - A_2L^2 - \dots - A_pL^p$ is a log polynomial of order p and u_t is a vector of iid error terms with zero mean and a fix variance-covariance matrix Σ_u . The

¹⁸I believe it is reasonable to assume that the index by [Baker et al. \(2014\)](#), which is based on an aggregation of news paper articles, is exogenous to my measure of disagreement extracted from the SPF. Therefore, a causal interpretation is valid at this point.

¹⁹[Andrade et al. \(2013\)](#) suggest to embed models of disagreement in a general equilibrium setup. In such a richer model it would be possible to test the model's predictions about the reaction of disagreement to different structural shocks.

optimal lag order is determined based on the Akaike information criterion and found to be $p = 2$.

To identify a structural VAR underlying the observed reduced-form representation several additional restrictions have to be made. I follow much of the standard macro literature and use the sign-restrictions approach proposed by Uhlig (2005) and Mountford and Uhlig (2009). With this approach identification is achieved by restricting the sign of the response of certain variables to a specific structural shock. Importantly, the sign of the response of the variable(s) of particular interest is left unrestricted a priori and is fully determined by the data. I identify an aggregate supply shock, an aggregate demand shock and a monetary policy shock. It is important to choose the imposed sign-restrictions in such a way that the different shocks are clearly identified. I do this by assuming that following an adverse aggregate supply shock gdp_t goes down and cpi_t as well as i_t go up, following an adverse aggregate demand shock gdp_t , cpi_t and i_t go down, and following a contractionary monetary policy shock i_t increases and gdp_t as well as cpi_t go down. These restrictions are imposed for the period of the shock and the following period, thus for two periods in total.²⁰

The impulse response figures (IRFs) for both measures of multivariate disagreement (based on short-term and based on long-term forecasts) to the three different shocks are shown in Figure 5.²¹ Two observations are noteworthy. First, the impact of each of the structural shocks tends to be larger on disagreement about the short-run outlook for the state of the economy relative to disagreement about the long-run prospects. Second, the only statistically significant responses (on a 90% confidence level) are those of short-term disagreement on aggregate demand shocks and of long-term disagreement on aggregate supply shocks. In other words: A negative demand shock leads to a significant (short-lived) increase in disagreement among forecasters about the state of the economy one year ahead, while an adverse aggregate supply shock leads to a significant (and persistent) rise in disagreement among forecasters about the state of the economy five years ahead. This seems plausible given that supply-side factors are most important for the long-term development of an economy while demand-side factors are more relevant for determining the state of the business-cycle in the short-run.

²⁰Results are qualitatively robust against varying this choice within a sensible range.

²¹I use the “pure” sign-restriction approach from Uhlig (2005). I draw up to 4.000 draws from the posterior of the parameters of the reduced-form VAR model. For each of these draws I simulate 100 potential impulse vectors. All impulse response functions are based on 4.000 draws that fulfill the requirements of the imposed sign-restrictions.

[Figure 5 about here.]

5 Implications for Models of Disagreement

Some of the stylized facts, which are documented in the previous sections, are not explained by existing models of forecast disagreement. Notably, I am referring to the persistence of forecasters' individual levels of relative disagreement and the low level of agreement on the basic structure of the economy. This suggests that future models should include additional features to provide a more realistic description of the process of heterogeneous expectation formation.

To explain in more detail how and why the observed data are at odds with the predictions of the current theoretical models, I first take a closer look at the model proposed by [Andrade et al. \(2013\)](#) as their model offers a multivariate framework to organize thoughts. This model of expectation formation assumes that the true data generating process for $y_t = [y_t^1, y_t^2, \dots, y_t^k]'$ is of the following form

$$y_t = (I_k - \Phi)\mu_t + \Phi y_{t-1} + \nu_t^y, \quad (5)$$

$$\mu_t = \mu_{t-1} + \nu_t^\mu, \quad (6)$$

with initial conditions y_0 and μ_0 . [Andrade et al. \(2013\)](#) assume furthermore that all eigenvalues of Φ are inside the unit circle and that the two error terms are *i.i.d.* Gaussian innovations with covariance matrices Σ^y and Σ^μ . Forecasters observe polluted idiosyncratic signals which are given by

$$y_{it} = y_t + \eta_{it}, \quad (7)$$

where η_t are *i.i.d.* Gaussian innovations with covariance matrix Σ^η . Forecasters then use the current and past values of y_{it} to construct optimal h-steps-ahead forecasts $\mathbb{E}[y_{t+h}|y_{it}, y_{i(t-1)}, \dots, y_{i0}]$ using the Kalman filter.²² Forecast disagreement occurs because forecasters are informationally constrained in the sense that they, first, do not perfectly observe the current true state of the economy and that they, second, have to infer what changes are due to transitory shocks (ν_t^y) and what changes are due to permanent shocks (ν_t^μ).

²²Details are given by equations 3.6-3.9 in [Andrade et al. \(2013\)](#).

The model is able to account for three features of forecast disagreement: i) disagreement is observed for all forecast horizons, ii) the shape of the term structure of disagreement may differ across different variables,²³ and iii) forecast disagreement is time-varying for all forecast horizons. The model fails, however, to generate predictions that are in line with the two additional stylized facts that have been documented in this paper: persistence in the ranking of relative individual disagreement and stark disagreement about the interaction of different variables.

Given the baseline parameter calibration from [Andrade et al. \(2013\)](#), simulating their model yields cross-sectional forecast correlations (implied by the off diagonal elements of $S_{t+h|t}$) that lie in the range of .32 to .89 in absolute values. This is much higher than the observed values which were documented in [Table 2](#). At the same time, when the simulated cross-sections of vector-forecasts are used to rank forecasters according to their individual level of relative disagreement for each time period, then Kendall's coefficient of concordance is, not surprisingly, close to 0. This is significantly lower than the observed value. In what follows I suggest ways to modify the model, which would enable it to account for these facts.

The first contradiction could be resolved either by increasing the variance of the noise component or by relaxing the assumption that all agents know the true structure of the model. The first solution seems not very attractive as it would increase the volatility of forecasts above the observed level and would lead to unrealistic signal-to-noise ratios.²⁴ The second solution could be implemented by introducing some kind of learning algorithm that agents use to infer the structure of the model based on the history of (polluted) data that they observe. This additional source of heterogeneity across agents would introduce more disagreement about the structure of dependence between different variables.

The second contradiction could be resolved by introducing heterogeneous signal-to-noise ratios that are agent-specific and fixed over time.²⁵ In this case, forecasters with

²³I do not discuss this issue in this paper because the term structure does not differ much across the three variables available in the SPF data set. [Andrade et al. \(2013\)](#) document that the term structure looks remarkable different for disagreement about future interest rates. This can be explained by the fact that interest rates, in contrast to most other macroeconomic variables, are observed without measurement error.

²⁴It could even be argued that the signal-to-noise ratios implied by the baseline calibration of [Andrade et al. \(2013\)](#) are already too low. [Coibion and Gorodnichenko \(2010\)](#), for instance, report estimated signal-to-noise ratios that are much higher.

²⁵[Coibion and Gorodnichenko \(2012\)](#) analyze this feature as an extension of their model. They reject it empirically based on indirect evidence that disagreement does not respond to shocks.

low signal-to-noise ratios would tend to deviate strongly from the average forecast, i.e., they would have persistently high levels of relative disagreement, while forecasters with high signal-to-noise ratios would tend to publish forecasts very similar to the average forecast, i.e., they would have persistently low levels of relative disagreement. A slight modification of the simulation of the model of [Andrade et al. \(2013\)](#) demonstrates this point: Instead of using a fixed Σ^n for all forecasters I multiply the Σ^n from their baseline calibration by a forecaster-specific factor, which is drawn randomly from a uniform distribution with a lower bound of .1 and an upper bound of 2. Simulating this modified model yields a coefficient of concordance for the sequence of rankings of relative disagreement of about .2, which is much closer to the values that are implied by the SPF data. Given that differences in the signal-to-noise ratio—which could reflect differences in the availability of private information, in the quality of the forecast production process, or different ways to process public signals—are very likely to exist in reality, this seems to be a plausible feature that could be added in a straightforward manner.

The alternative type of model as suggested by [Lahiri and Sheng \(2008\)](#) or [Patton and Timmermann \(2010\)](#) has been proposed to model sequences of fixed-event forecasts. The prior about the long-run outlook is re-drawn from a particular distribution for each of these events (normally annual growth rates). [Lahiri and Sheng \(2008\)](#) assume, for instance, that each January every forecaster forms some prior about the annual growth rate in the next calendar year. They do not suggest any dependence between the draws for two consecutive years. This would, however, be necessary to generate a certain degree of persistence in the ranking of levels of individual relative disagreement. In the context of fixed-horizon forecasts, one possibility (following [Coibion and Gorodnichenko \(2012\)](#)) is to make the ad-hoc assumption that forecasters build their forecasts as a weighted average of the optimal model-consistent forecast and some agent-specific prior, which is assumed to be fixed over time. The weights given to the prior at different forecast horizons could be calibrated to match the observed persistence of the ranking. Such a model could also generate the observed persistence of the ranking when forecasters are ranked according to their point predictions. This approach seems, however, not very attractive since forecasts are formed in an ad-hoc way and since the assumption of a strictly fixed long-run prior appears overly strict.

6 Concluding Remarks

This paper analyzes the behavior of disagreement among professional forecasters of the economy in the Euro area over the period from 1999 to 2014. It is the first documentation of disagreement among professional forecasters from a multivariate perspective. I have used measures proposed by [Banerghansa and McCracken \(2009\)](#) that take into account both dispersion of univariate forecasts as well as disagreement about the co-movement of variables.

I have documented various stylized facts of multivariate disagreement and I have discussed what implications they have for the theoretical modelling of heterogeneous expectation formation.

I have shown that overall disagreement fluctuates substantially over time. These fluctuations are mainly driven by changes of disagreement about the future of individual variables rather than changes in the extent to which forecasters agree or disagree about the expected correlation between the different variables. In that sense, it is the volatility of disagreement about scenarios rather than variations of the degree of disagreement about the structure of the economy that is driving movements of multivariate disagreement. However, looking at the unconditional level of multivariate disagreement reveals that disagreement about the structure of the economy contributes substantially to overall disagreement.

Furthermore, I have shown that there is some degree of rank persistence over time when forecasters are ranked according to their individual level of relative disagreement. My results indicate also that variations of overall disagreement are driven by economic uncertainty. More precisely, policy uncertainty affects disagreement about the long-run outlook while proxies that measure uncertainty on financial markets are more closely related to disagreement about the economic situation in the short-run. Finally, I have shown that disagreement about the long-run is affected persistently and significantly by aggregate supply shocks while disagreement about the near future is temporarily and significantly affected by aggregate demand shocks.

The last part of the paper shows that existing models of forecast disagreement, by construction, do not capture important aspects of heterogeneity across forecasters. In particular they do not allow for much disagreement about the structure of the economy and predict no persistence in the level of individual relative disagreement. A promising approach to reconcile theoretical models of heterogeneous expectation formation and these documented stylized facts about disagreement seems to be a modified version of

the model by [Andrade et al. \(2013\)](#), in which forecasters face different signal-to-noise ratios (to generate both consensus-oriented and disagreeing types of forecasters) and have to learn about the parameters of the model (to generate less agreement about the structure of the economy). Using information about multivariate aspects of disagreement can be beneficial for the calibration of such models.

References

- Andrade, Philippe and Hervé Le Bihan**, “Inattentive professional forecasters,” *Journal of Monetary Economics*, 2013, 60 (8), 967–982.
- , **Richard K. Crump, Stefano Eusepi, and Emanuel Moench**, “Noisy Information and Fundamental Disagreement,” Staff Report 655, Federal Reserve Bank of New York 2013.
- Bachmann, Rüdiger, Steffen Elstner, and Eric R. Sims**, “Uncertainty and Economic Activity: Evidence from Business Survey Data,” *American Economic Journal: Macroeconomics*, 2013, 5 (2), 217–49.
- Baker, Scott, Nicholas Bloom, and Steven J. Davis**, “Measuring Economic Policy Uncertainty,” <http://www.policyuncertainty.com> 2014. Online; accessed 2014-07-21.
- Ball, Laurence M., João Tovar Jalles, and Prakash Loungani**, “Do Forecasters Believe in Okun’s Law? An Assessment of Unemployment and Output Forecasts,” IMF Working Papers 14/24, International Monetary Fund 2014.
- Banternghansa, Chanont and Michael W. McCracken**, “Forecast disagreement among FOMC members,” Working Papers 2009-059, Federal Reserve Bank of St. Louis 2009.
- Batchelor, Roy**, “Bias in macroeconomic forecasts,” *International Journal of Forecasting*, 2007, 23, 189–203.
- Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis**, “The measurement and characteristics of professional forecasters’ uncertainty,” *Journal of Applied Econometrics*, 2014, *forthcoming*.
- Bruine De Bruin, Wändi, Charles F. Manski, Giorgio Topa, and Wilbert van der Klaauw**, “Measuring consumer uncertainty about future inflation,” *Journal of Applied Econometrics*, 2011, 26 (3), 454–478.
- Carroll, Christopher D.**, “Macroeconomic Expectations of Households and Professional Forecasters,” *Quarterly Journal of Economics*, 2003, 118 (1), 269–298.
- Coibion, Olivier and Yuriy Gorodnichenko**, “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts,” NBER Working Papers 16537, National Bureau of Economic Research 2010.
- and —, “What Can Survey Forecasts Tell Us about Information Rigidities?,” *Journal of Political Economy*, 2012, 120 (1), 116 – 159.
- Dovern, Jonas**, “When are GDP forecasts updated? Evidence from a large international panel,” *Economics Letters*, 2013, 120 (3), 521–524.
- , **Ulrich Fritsche, and Jiri Slacalek**, “Disagreement Among Forecasters in G7 Countries,” *The Review of Economics and Statistics*, 2012, 94 (4), 1081–1096.
- , —, **Prakash Loungani, and Natalia T. Tamirisa**, “Information Rigidities: Comparing Average and Individual Forecasts for a Large International Panel,” *International Journal of Forecasting*, 2014, *forthcoming*.
- Fendel, Ralf, Eliza M. Lis, and Jan-Christoph Rülke**, “Do professional forecasters believe in the Phillips curve? evidence from the G7 countries,” *Journal of Forecasting*, 2011, 30 (2), 268–287.

- Frenkel, Michael, Eliza M. Lis, and Jan-Christoph Rülke**, “Has the economic crisis of 2007-2009 changed the expectation formation process in the Euro area?,” *Economic Modelling*, 2011, 28 (4), 1808–1814.
- Giordani, Paolo and Paul Söderlind**, “Inflation forecast uncertainty,” *European Economic Review*, 2003, 47, 1037–1059.
- Kendall, Maurice G.**, *Rank Correlation Methods*, Charles Griffin & Company, 1948.
- Lahiri, Kajal and Xuguang Sheng**, “Evolution of Forecast Disagreement in a Bayesian Learning Model,” *Journal of Econometrics*, 2008, 144 (2), 325–340.
- and —, “Measuring forecast uncertainty by disagreement: The missing link,” *Journal of Applied Econometrics*, 2010, 25 (4), 514–538.
- Mankiw, N. Gregory and Ricardo Reis**, “Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *Quarterly Journal of Economics*, 2002, 117, 1295–1328.
- , —, and **Justin Wolfers**, “Disagreement on Inflation Expectations,” *NBER Macroeconomics Annual*, 2003, pp. 209–248.
- Mountford, Andrew and Harald Uhlig**, “What are the effects of fiscal policy shocks?,” *Journal of Applied Econometrics*, 2009, 24 (6), 960–992.
- Patton, Andrew J. and Allan Timmermann**, “Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion,” *Journal of Monetary Economics*, October 2010, 57 (7), 803–820.
- Rich, Robert and Joseph Tracy**, “The Relationships among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts,” *The Review of Economics and Statistics*, 2010, 92 (1), 200–207.
- Sims, Christopher**, “Implications of Rational Inattention,” *Journal of Monetary Economics*, 2003, 50 (3), 665–690.
- Uhlig, Harald**, “What are the effects of monetary policy on output? Results from an agnostic identification procedure,” *Journal of Monetary Economics*, 2005, 52 (2), 381–419.
- Woodford, Michael**, “Imperfect Common Knowledge and the Effects of Monetary Policy,” in Philippe Aghion, Roman Frydman Joseph Stiglitz, and Michael Woodford, eds., *Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, Princeton University Press Princeton 2003, pp. 25–58.

Table 1: Descriptive Statistics for SPF Data

	Growth (g)		Inflation (π)		Unempl. (u)	
	$h = 4$	$h = 20$	$h = 4$	$h = 20$	$h = 4$	$h = 20$
N	2447	1874	2447	1874	2447	1874
N/T	40.78	34.70	40.78	34.70	40.78	34.70
SD(N/T)	5.52	4.00	5.52	4.00	5.52	4.00
Average	1.56	2.04	1.75	1.90	9.12	8.03
SD	1.06	0.37	0.37	0.24	1.37	1.08

Notes: N denotes the total number of observations; N/T is the average number of observations per survey wave; SD(N/T) is the standard deviation over time of this average number of observations; Average denotes the average value of the point forecasts; SD denotes the standard deviation of point forecasts (across individuals and time).

Table 2: Average Values of Elements of $S_{t+h|t}$

	$h = 4$	$h = 20$
$sd(g)$	0.36	0.27
$sd(\pi)$	0.27	0.22
$sd(u)$	0.31	0.65
$Corr(g, \pi)$	0.16	0.10
$Corr(g, u)$	-0.25	-0.17
$Corr(\pi, u)$	-0.10	-0.02

Notes: The values shown are based on the averages for each element of $S_{t+h|t}$ over the full sample. These averages are computed as $\bar{s}_{ij,h} = \sum_{t=1}^T s_{t+h|t}^{ij}$, where $s_{t+h|t}^{ij}$ denotes the element of $S_{t+h|t}$ corresponding to the i^{th} row and the j^{th} column. $sd(\bullet)$ is computed as the square root of the main diagonal elements. $Corr(\bullet, \bullet)$ is computed as the correlations that correspond to the off-diagonal elements.

Table 3: Changes in Multivariate Disagreement after the Great Recession

	Short-Term ($h = 4$)						
	$D_{t+4 t}$	$sd(g)$	$sd(\pi)$	$sd(u)$	$Corr(g, \pi)$	$Corr(g, u)$	$Corr(\pi, u)$
Crisis	0.078*** (0.019)	0.190*** (0.060)	0.099*** (0.027)	0.098** (0.037)	0.044 (0.074)	0.110 (0.085)	0.167** (0.073)
Post 2008q3	0.015 (0.013)	0.066 (0.042)	0.047** (0.019)	0.084*** (0.026)	-0.014 (0.052)	-0.081 (0.060)	-0.034 (0.052)
Constant	0.020*** (0.007)	0.307*** (0.022)	0.243*** (0.010)	0.268*** (0.014)	-0.252*** (0.027)	0.173*** (0.031)	-0.111*** (0.027)
N	60	60	60	60	60	60	60
R^2	0.35	0.30	0.41	0.38	0.01	0.04	0.09
	Long-Term ($h = 20$)						
	$D_{t+20 t}$	$sd(g)$	$sd(\pi)$	$sd(u)$	$Corr(g, \pi)$	$Corr(g, u)$	$Corr(\pi, u)$
Crisis	-0.000 (0.010)	0.043* (0.022)	-0.035 (0.026)	-0.057 (0.065)	0.134* (0.073)	0.176** (0.077)	0.086 (0.074)
Post 2008q3	0.024*** (0.007)	0.003 (0.016)	0.063*** (0.019)	0.220*** (0.047)	-0.026 (0.053)	-0.056 (0.056)	-0.071 (0.053)
Constant	0.030*** (0.004)	0.261*** (0.009)	0.203*** (0.010)	0.570*** (0.026)	-0.181*** (0.029)	0.097*** (0.031)	-0.008 (0.029)
N	54	54	54	54	54	54	54
R^2	0.21	0.10	0.18	0.32	0.07	0.09	0.04

Notes: $D_{t+h|t}$ is the measure for multivariate disagreement given by equation 2. $sd(\bullet)$ refers to the cross-sectional standard deviation for a particular type of univariate forecast. $Corr(\bullet, \bullet)$ refers to the cross-sectional correlation of a particular pair of univariate forecasts. *Crisis* is a dummy variable that is equal to 1 for the periods from 2008q3 to 2010q2 and 0 otherwise. *Post2008q3* is a dummy variable that is 1 since 2008q3 and 0 for all periods before. Standard errors are given in parenthesis. *, ** and *** denote significance of coefficients on the 10%, 5% and 1% confidence level.

Table 4: Persistence of Ranking With Respect to Level of Disagreement

	$h = 4$	$h = 20$
Full sample	0.14	0.20
Before 2008q3	0.16	0.31
Since 2008q3	0.25	0.32
<i>For comparison (full sample):</i>		
Point forecasts		
g	.13	.30
π	.19	.44
u	.14	.27
Forecast uncertainty		
g	.56	.65
π	.32	.62
u	.54	.56

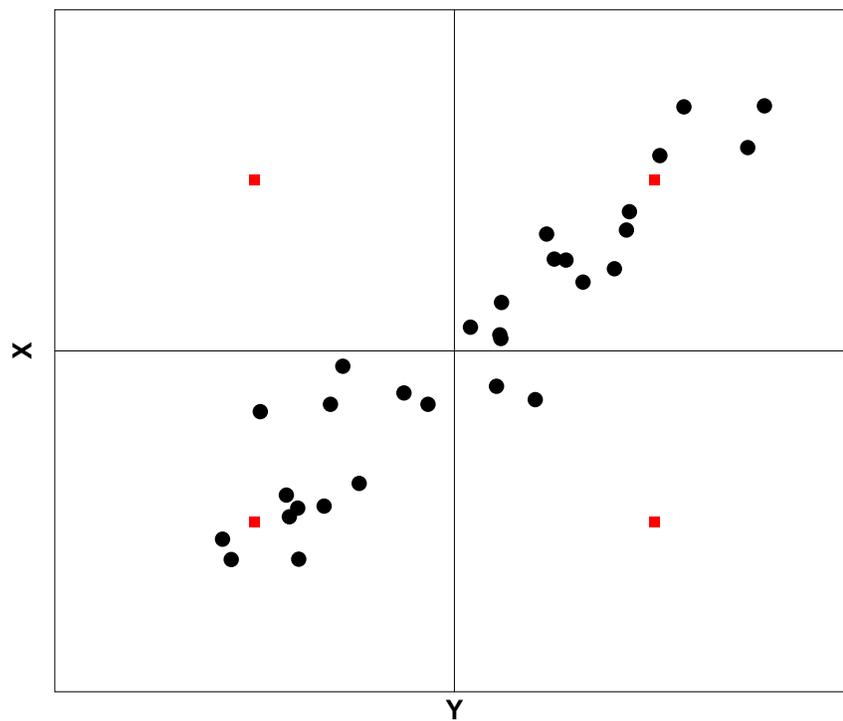
Notes: The table shows values of Kendall's coefficient of concordance. The statistics have been calculated by accounting for missing values as proposed in [Boero et al. \(2014\)](#).

Table 5: Correlation between Multivariate Disagreement and Measures of Economic Uncertainty

	$h = 4$					$h = 20$				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
Constant	-0.002 (0.014)	-0.017 (0.013)	-0.048** (0.014)	-0.070*** (0.018)	-0.086*** (0.020)	0.005 (0.009)	0.010 (0.009)	0.004 (0.010)	-0.010 (0.012)	-0.022* (0.012)
$D_{t-1+h} _{t-1}$	0.535*** (0.110)	0.500*** (0.105)	0.479*** (0.093)	0.467*** (0.094)	0.445*** (0.091)	0.224* (0.131)	0.338** (0.127)	0.378*** (0.127)	0.299** (0.121)	0.210 (0.127)
$Bloom_{t-1}$	0.014 (0.010)				0.016 (0.010)	0.018*** (0.006)				0.017** (0.007)
$ESTX50V_{t-1}$		0.002*** (0.001)			-0.001 (0.001)		0.001** (0.000)			-0.000 (0.000)
OPV_{t-1}			0.002*** (0.000)		0.002*** (0.001)			0.001** (0.000)		0.000 (0.000)
FXV_{t-1}				0.009*** (0.002)	0.004 (0.003)				0.004*** (0.001)	0.003 (0.002)
N	59	59	59	59	59	51	51	51	51	51
R^2	0.33	0.41	0.53	0.52	0.59	0.26	0.22	0.22	0.30	0.38

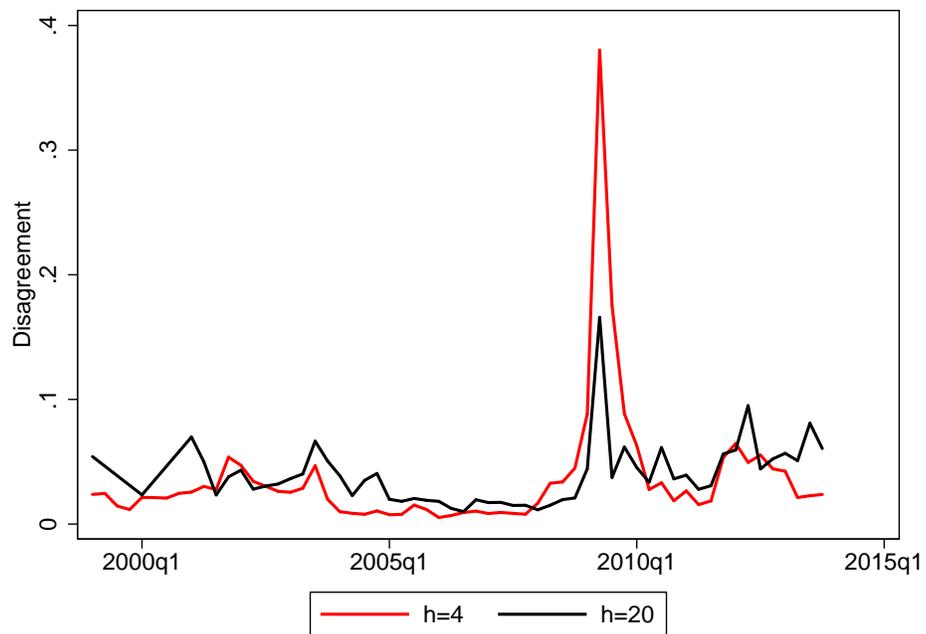
Notes: The dependent variable in each model is $D_{t+4|t}$ or $D_{t+20|t}$ respectively. $Bloom_{t-1}$ refers to the lagged policy uncertainty index provided by Baker et al. (2014). $ESTX50V_{t-1}$ refers to the lagged volatility of the EuroStoxx50 measured as the realized (annualized) volatility based on daily stock market returns. OPV_{t-1} refers to the lagged volatility of oil prices measured as the realized (annualized) volatility based on daily changes of the WTI oil price (in US-Dollar). FXV_{t-1} refers to the lagged volatility of the exchange rate of the Euro against the US-Dollar measured as the realized (annualized) volatility based on daily changes of the exchange rate. Standard errors are given in parenthesis. *, **, and *** denote significance of coefficients on the 10%, 5% and 1% confidence level.

Figure 1: Intuitive Argument for Using Multivariate Disagreement



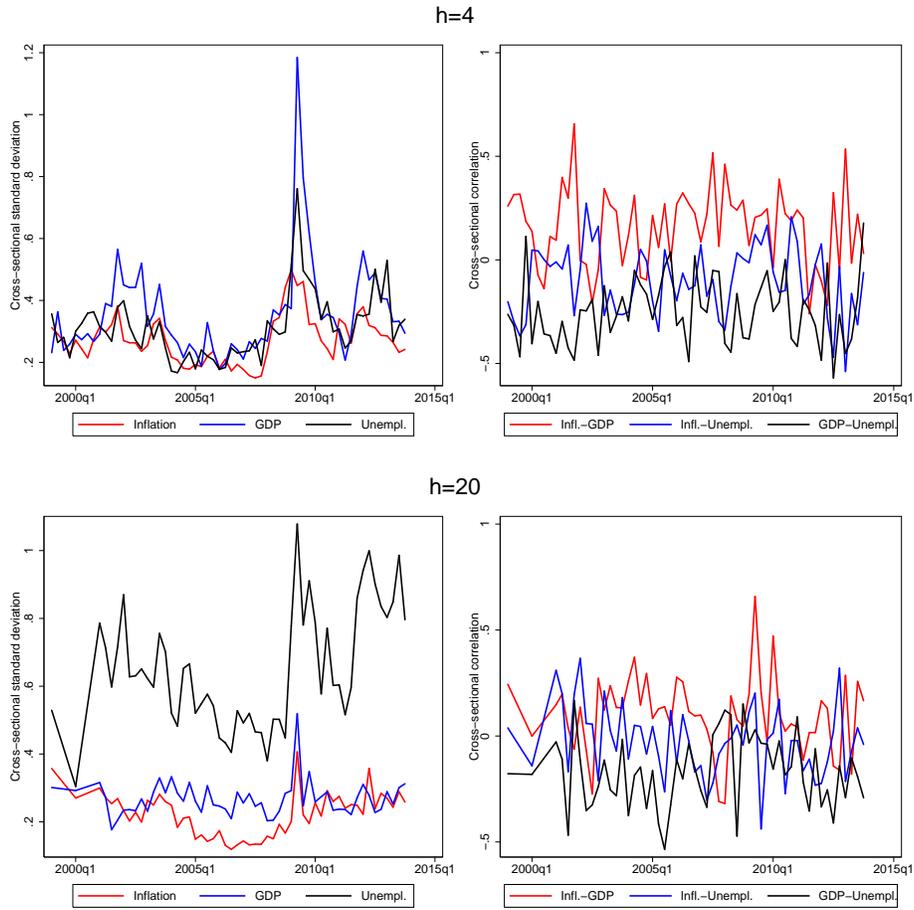
Notes: Data for the black dots were generated using a bivariate standard normal distribution with a correlation coefficient of 0.9. The figure illustrates that univariate measures of disagreement or multivariate measures that do not take correlations between variables into account (such as the Euclidean distance) miss important aspects when variables are correlated.

Figure 2: Multivariate Disagreement among SPF Participants



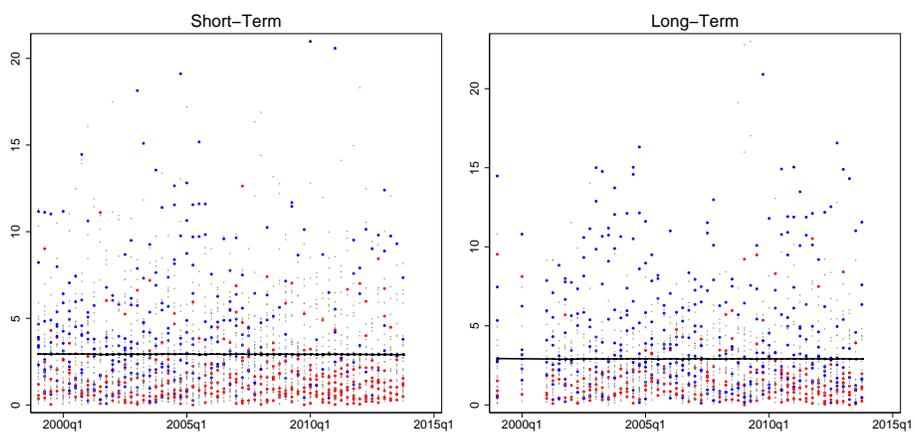
Notes: Estimates are computed using equation 2.

Figure 3: Elements of Multivariate Disagreement



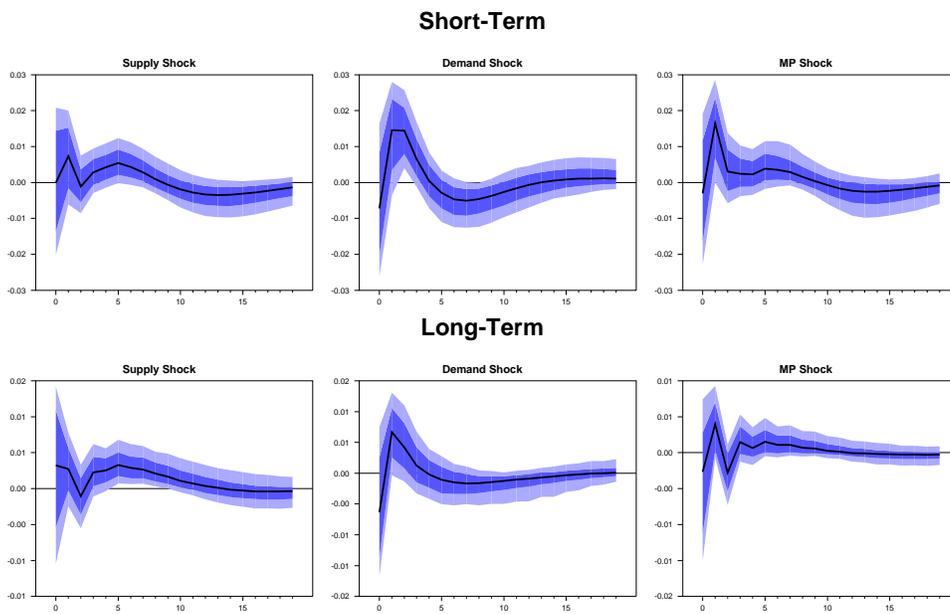
Notes: The left plots show the square roots of the elements from the main-diagonal of $S_{t+h|t}$ as a univariate measure of disagreement about the future paths of each individual variable of y_t . The right plots show the cross-sectional correlations implied by $S_{t+h|t}$ for the three possible bivariate combinations of the variables in y_t .

Figure 4: Persistence of Individual Levels of Disagreement



Notes: Individual level of disagreement computed as in equation 1 is measured on the vertical axis. The blue dots refer to observations corresponding to those 10 forecasters that have the highest average rank in terms of disagreement over the entire sample. The red dots refer to observations corresponding to those 10 forecasters that have the lowest average rank in terms of disagreement over the entire sample. The small grey dots refer to all other observations. The black line represents the average level of disagreement at each point in time and has been added to demonstrate that $d_{i,t+h|t}$ is a normalized measure.

Figure 5: Reaction of Disagreement to Structural Shocks



Notes: The upper plots show the impulse response functions of $D_{t+4|t}$ based on a VAR given by equation 4. The lower plots show the distribution of IRFs of $D_{t+20|t}$ based on a VAR given by equation 4. The black lines represent the median response while the dark (light) blue bands represent 66% (90%) confidence bands based on 4,000 simulated IRFs.