

## MEASURING FORECAST UNCERTAINTY BY DISAGREEMENT: THE MISSING LINK

KAJAL LAHIRI<sup>a\*</sup> AND XUGUANG SHENG<sup>b</sup>

<sup>a</sup> *Department of Economics, University at Albany, SUNY, Albany, NY, USA*

<sup>b</sup> *Department of Economics, American University, Washington, D. C., USA*

### SUMMARY

Using a standard decomposition of forecast errors into common and idiosyncratic shocks, we show that aggregate forecast uncertainty can be expressed as the disagreement among the forecasters plus the perceived variability of future aggregate shocks. Thus the reliability of disagreement as a proxy for uncertainty will be determined by the stability of the forecasting environment and the length of the forecast horizon. Using density forecasts from the Survey of Professional Forecasters, we find direct evidence in support of our hypothesis. Our results support the use of GARCH-type models, rather than the *ex post* squared errors in consensus forecasts, to estimate the *ex ante* variability of aggregate shocks as a component of aggregate uncertainty. Copyright © 2010 John Wiley & Sons, Ltd.

*Received 31 March 2008; Revised 15 October 2008*

### 1. INTRODUCTION

Forecast uncertainty is playing an increasingly important role in macroeconomics and monetary policy making. Since the mid 1990s, the Bank of England and Sveriges Riksbank have been reporting fan charts that show subjective confidence bands surrounding official forecasts. Effective November 2007, each member of the US Federal Open Market Committee (FOMC) is also publishing information about uncertainty associated with their economic outlooks. These advances in the quantification and communication strategies by central banks are expected to contribute to a more informed discussion and monitoring of future economic prospects than is possible with point forecasts alone (see Bernanke, 2007; Wallis, 2004).

Since forecast uncertainty is intrinsically unobservable, evaluating its estimates poses challenging methodological problems. As a result, economists have experimented with alternative proxies for forecast uncertainty. One of the more popular real-time measures has been forecast disagreement, simply calculated as the dispersion in alternative point forecasts. When disagreement is taken to indicate uncertainty, the underlying assumption is that this inter-personal dispersion measure is an acceptable proxy for the average dispersion of intra-personal predictive probabilities held by individual experts. The validity of this assumption can by no means be taken for granted. Since the seminal work of Zarnowitz and Lambros (1987), economists have studied but disagreed on whether disagreement is a good proxy for uncertainty.<sup>1</sup> As pointed out by Bomberger (1996) and

---

\* Correspondence to: Kajal Lahiri, Department of Economics, University at Albany, SUNY, 1400 Washington Av., Albany, NY 12222, USA. E-mail: klahiri@albany.edu

<sup>1</sup> See, for instance, Bomberger (1996, 1999), Rich and Butler (1998), Giordani and Söderlind (2003), Lahiri and Liu (2005), and Boreo *et al.* (2008).

Giordani and Söderlind (2003), disagreement remains a theoretically unfounded measure of uncertainty. Interestingly, there has been parallel but largely independent research in the accounting and finance literature on whether disagreement among financial or market analysts can be used as a proxy for uncertainty about future earnings.<sup>2</sup>

In this paper, we establish a simple relationship connecting forecast uncertainty to disagreement. Using a standard decomposition of forecast errors into common and idiosyncratic components, we show that forecast uncertainty equals disagreement plus the variance of future aggregate shocks that accumulate over the horizons. This finding has important implications for the empirical studies that use disagreement as a proxy for uncertainty. It suggests that the robustness of the proxy depends on the variance of aggregate shocks over time and across horizons. It also simplifies the multidimensional covariance matrix of individual forecast errors in Barry and Jennings (1992) in terms of the variance of aggregate shocks, which can be easily interpreted as the uncertainty shared by all forecasters due to their exposure to future common shocks.

Using a panel of density forecasts from the Survey of Professional Forecasters over 1969–2007, we find direct evidence in support of our hypothesized time and horizon effects. As for the time effect, disagreement is found to be a reliable measure for uncertainty in stable periods. In periods with large volatility of aggregate shocks; however, disagreement becomes a less reliable proxy. As for the horizon effect, we find that the longer the forecast horizon, the larger is the difference between disagreement and uncertainty.

In recent accounting and finance literature, squared errors in consensus forecasts have been used as proxies for the variance of future aggregate shocks as a component of forecast uncertainty. Our results suggest that adding the squared mean forecast error to disagreement can make the estimated uncertainty worse than using disagreement alone. If one wants to construct a robust *ex ante* measure of uncertainty, our suggestion is to use the sum of the observed disagreements from the survey and the variance of future aggregate shocks generated by GARCH-type models that use a moving average squared error over the past few years as one of the covariates.

The remainder of the paper is organized as follows. In Section 2 we develop the theoretical model and derive the relationship between disagreement and uncertainty. Section 3 tests empirically whether disagreement is a reliable proxy for uncertainty and suggests a method to construct the *ex ante* measure of uncertainty. Section 4 concludes.

## 2. THE ECONOMETRIC MODEL

This section begins with deriving the unconditional relationship between forecast uncertainty and disagreement. In this context, we then connect our framework to the Bayesian learning model. The conditional relationship between forecast uncertainty and disagreement that is relevant in real time is presented at the end of this section.

### 2.1. The Unconditional Relationship between Uncertainty and Disagreement

For  $N$  individuals,  $T$  target years,  $H$  forecast horizons, let  $F_{ith}$  be the forecast of the variable of interest made by agent  $i$ , for the target year  $t$  and  $h$  quarters ahead to the end of the target

---

<sup>2</sup> See Zhang (2006) and references therein. Barry and Jennings (1992), Abarbanell *et al.* (1995), Barron *et al.* (1998), Diether *et al.* (2002) and Johnson (2004) have argued that disagreement alone is not sufficient to approximate uncertainty.

year, and  $A_t$  be the actual value of variable. The individual forecast error ( $e_{ith}$ ) is defined as

$$e_{ith} = A_t - F_{ith} \quad (1)$$

Following Davies and Lahiri (1995, 1999), we write  $e_{ith}$  as the sum of a component common to all forecasters ( $\lambda_{th}$ ) and idiosyncratic errors ( $\varepsilon_{ith}$ ):

$$e_{ith} = \lambda_{th} + \varepsilon_{ith} \quad (2)$$

$$\lambda_{th} = \sum_{j=1}^h u_{tj} \quad (3)$$

The common component ( $\lambda_{th}$ ) represents the cumulative effect of all shocks that occurred from  $h$  quarters ahead to the end of target year  $t$ . Equation (3) shows that this accumulation of shocks is the sum of each quarterly shock ( $u_{tj}$ ) that occurred over the span. Even if forecasters make ‘perfect’ forecasts, the forecast error may still be nonzero due to shocks which are, by nature, unpredictable. Forecasters, however, do not make ‘perfect’ forecasts even in the absence of unanticipated shocks. This ‘lack of perfection’ is due to other factors (e.g., differences in information acquisition and processing, interpretation, judgment, and forecasting models) specific to a given individual at a given point in time and is represented by the idiosyncratic error ( $\varepsilon_{ith}$ ).

We make the following simplifying assumptions:

**Assumption 1**  $E(u_{tj}) = 0$ ;  $\text{var}(u_{tj}) = \sigma_{u|tj}^2$  for any  $t$  and  $j$ ;  $E(u_{tj}u_{ts}) = 0$  for any  $t$  and  $j \neq s$ ;  $E(u_{th}u_{t-k,h}) = 0$  for any  $t$ ,  $h$  and  $k \neq 0$ .

**Assumption 2**  $E(\varepsilon_{ith}) = 0$ ;  $\text{var}(\varepsilon_{ith}) = \sigma_{\varepsilon_{ith}}^2$  for any  $i$ ,  $t$  and  $h$ ;  $E(\varepsilon_{ith}\varepsilon_{jth}) = 0$  for any  $t$ ,  $h$  and  $i \neq j$ .

**Assumption 3**  $E(\varepsilon_{ith}u_{t-k,j}) = 0$  for any  $i$ ,  $t$ ,  $h$ ,  $k$  and  $j$ .

Thus aggregate shocks are assumed to be uncorrelated over time and horizons (Assumption 1). The idiosyncratic errors are taken to be mutually uncorrelated at all leads and lags (Assumption 2). In addition, the common component and idiosyncratic disturbances are assumed to be uncorrelated at all leads and lags (Assumption 3), which is a standard assumption in the literature. Taken together, Assumptions 1–3 imply that the individual forecast error has the factor model interpretation.

The observed disagreement ( $d_{th}$ ) among forecasters is the variance of their point forecasts which, given (1) and (2), can be expressed as

$$d_{th} \equiv \frac{1}{N-1} \sum_{i=1}^N (F_{ith} - F_{.th})^2 = \frac{1}{N-1} \sum_{i=1}^N (\varepsilon_{ith} - \frac{1}{N} \sum_{j=1}^N \varepsilon_{jth})^2 \quad (4)$$

where  $F_{.ith} = \frac{1}{N} \sum_{j=1}^N F_{jith}$ .<sup>3</sup> Note that the sample variance  $d_{ith}$  is a random variable prior to observing forecasts. Taking expectations, we obtain an expression for the non-random disagreement, denoted by  $D_{ith}$ , as

$$\begin{aligned} D_{ith} \equiv E(d_{ith}) &= \frac{1}{N-1} \sum_{i=1}^N E\left(\varepsilon_{ith} - \frac{1}{N} \sum_{j=1}^N \varepsilon_{jith}\right)^2 \\ &= \frac{1}{N-1} \sum_{i=1}^N (\sigma_{\varepsilon|ith}^2 + \frac{1}{N^2} \sum_{j=1}^N \sigma_{\varepsilon|jith}^2 - \frac{2}{N} \sigma_{\varepsilon|ith}^2) \\ &= \frac{1}{N} \sum_{i=1}^N \sigma_{\varepsilon|ith}^2 \end{aligned} \quad (5)$$

Thus, not surprisingly, we find that  $D_{ith}$  is determined by the average variance of idiosyncratic errors.

The uncertainty associated with a forecast of any specific individual is measured by the variance of the individual forecast error, and can be expressed as

$$U_{ith} = \text{var}(A_t - F_{ith}) = \text{var}(\lambda_{th} + \varepsilon_{ith}) = \sigma_{\lambda|th}^2 + \sigma_{\varepsilon|ith}^2 \quad (6)$$

Individual forecast uncertainty in (6) is comprised of two components: perceived uncertainty associated with forthcoming common shocks  $\sigma_{\lambda|th}^2$  and the variance of idiosyncratic shocks  $\sigma_{\varepsilon|ith}^2$ . Following Zarnowitz and Lambros (1987), we measure overall forecast uncertainty ( $U_{ith}$ ) as the average of the individual forecast error variances  $U_{ith} \equiv \frac{1}{N} \sum_{i=1}^N U_{ith}$ , which can be interpreted as the confidence an outside observer will have in a randomly drawn typical individual forecast from the panel of forecasters.<sup>4</sup> Given our model,  $U_{ith}$  can be expressed as a function of the model parameters as

$$U_{ith} = \sigma_{\lambda|th}^2 + \frac{1}{N} \sum_{i=1}^N \sigma_{\varepsilon|ith}^2 \quad (7)$$

After substituting (5) into (7), we get

$$U_{ith} = \sigma_{\lambda|th}^2 + D_{ith} \quad (8)$$

Given the model assumptions, forecast uncertainty, disagreement and the variance of forthcoming aggregate shocks are expected to be related in the sample as in (8): uncertainty is simply the disagreement plus the variance of the accumulated aggregate shocks over the forecast horizon. Thus the wedge between uncertainty and disagreement will be determined partly by the size of the

<sup>3</sup> The number of forecasters in the survey changes over both  $t$  and  $h$ . For simplicity, we suppress the subscripts  $t$  and  $h$  of  $N$  in equation (4) and thereafter.

<sup>4</sup> See also Lahiri *et al.* (1988), Batchelor and Dua (1996), Bomberger (1996), Giordani and Söderlind (2003), and Boero *et al.* (2008).

forecast horizon over which the aggregate shocks accumulate—the longer is the forecast horizon, the bigger will be the difference on average. It also suggests that the robustness of the relationship between the two will depend on the variability of aggregate shocks over time. In relatively stable time periods where the perceived variability of the aggregate shocks is small, whether the perceptions are correct or not, disagreement will be a good proxy for the unobservable aggregate uncertainty. In periods where the perceived volatility of the aggregate shocks is high, disagreement can become a tenuous proxy for uncertainty.

## 2.2. Connection to Bayesian Learning Model

In our current framework, we model the variance of forecast errors without modeling forecasters' expectation formation process. Actually, it is easy to connect our model with Bayesian learning framework that explicitly models individuals' forecasting behavior. Suppose that each forecaster is endowed with two signals: one public signal, represented by

$$l_{ith} = A_t + \eta_{ith}, \quad \eta_{ith} \sim N(0, 1/\sigma_{\eta|ith}^2) \quad (9)$$

and one private signal, represented by

$$s_{ith} = A_t + \zeta_{ith}, \quad \zeta_{ith} \sim N(0, 1/\sigma_{\zeta|ith}^2) \quad (10)$$

The private signal is assumed to be independent of the public signal and also independent of other private signals, which are standard assumptions in the literature (see Lahiri and Sheng, 2008). Each forecaster then combines these two sources of information, via Bayes' rule, to derive the conditional expected value of  $A_t$  as<sup>5</sup>

$$F_{ith} \equiv E(A_t | l_{ith}, s_{ith}) = (\sigma_{\eta|ith}^2 l_{ith} + \sigma_{\zeta|ith}^2 s_{ith}) / (\sigma_{\eta|ith}^2 + \sigma_{\zeta|ith}^2) \quad (11)$$

and the conditional variance of  $A_t$  as

$$U_{ith} \equiv \text{var}(A_t | l_{ith}, s_{ith}) = 1 / (\sigma_{\eta|ith}^2 + \sigma_{\zeta|ith}^2) \quad (12)$$

The individual forecast uncertainty defined in (12) reflects the uncertainty in both the public and private information, which is similar to (6), where the individual forecast uncertainty is comprised of perceived uncertainty associated with forthcoming common shocks and idiosyncratic shocks. Then we measure overall forecast uncertainty ( $U_{th}$ ) as the average of the individual uncertainties  $U_{ith} \equiv \frac{1}{N} \sum_{i=1}^N U_{ith}$ . Given the Bayesian learning model,  $U_{th}$  can be expressed as

$$U_{th} = \frac{1}{N} \sum_{i=1}^N \frac{1}{(\sigma_{\eta|ith}^2 + \sigma_{\zeta|ith}^2)} \quad (13)$$

<sup>5</sup> Under the assumption that  $\sigma_{\zeta|ith}^2 = \sigma_{\eta|ith}^2$  for all  $i$ , the individual forecast error can be written as  $e_{ith} = [-\sigma_{\eta|ith}^2 / (\sigma_{\eta|ith}^2 + \sigma_{\zeta|ith}^2)] \eta_{ith} + [-\sigma_{\zeta|ith}^2 / (\sigma_{\eta|ith}^2 + \sigma_{\zeta|ith}^2)] \zeta_{ith}$ , where the first and second terms on the right-hand side correspond to  $\lambda_{ith}$  and  $\varepsilon_{ith}$  in (2), respectively.

Note that overall forecast uncertainty in (13), derived in the context of a Bayesian learning framework, provides the justification that the aggregate uncertainty can be defined as the simple average of individual uncertainties as in (7). It is the combined uncertainty with equal weights in the context of the forecast combination literature.<sup>6</sup>

The disagreement among forecasters can be measured by the expected dispersion of  $F_{ith}$ . To examine the effect of new information on the disagreement, we consider the pre-posterior variance of opinions across forecasters. For any given information system represented by  $\sigma_{\eta|th}^2$  and  $\sigma_{\zeta|ith}^2$ , the pre-posterior variance is the variance based on the distribution of the signals  $l_{ith}$  and  $s_{ith}$  for  $i = 1, 2, \dots, N$ . The disagreement among forecasters can then be measured as

$$\begin{aligned}
 D_{th} &\equiv E \left[ \frac{1}{N-1} \sum_{i=1}^N (F_{ith} - \frac{1}{N} \sum_{j=1}^N F_{jth})^2 \right] \\
 &= \frac{1}{N} E \left[ \sum_{i=1}^N F_{ith}^2 \right] - \frac{1}{N(N-1)} E \left[ \sum_{i=1}^N \sum_{j \neq i}^N F_{ith} F_{jth} \right] \tag{14}
 \end{aligned}$$

After substituting for  $F_{ith}$  from (11), we get

$$D_{th} = \left[ \frac{1}{N} \sum_{i=1}^N \frac{1}{(\sigma_{\eta|th}^2 + \sigma_{\zeta|ith}^2)} \right] - \left[ \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N \frac{\sigma_{\eta|th}^2}{(\sigma_{\eta|th}^2 + \sigma_{\zeta|ith}^2)(\sigma_{\eta|th}^2 + \sigma_{\zeta|jth}^2)} \right] \tag{15}$$

Note that the first term on the right-hand side of (15) is forecast uncertainty,  $U_{th}$  and the second term is the average covariance among forecast errors,  $C_{th}$ , where

$$C_{th} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N \text{cov}(A_t - F_{ith}, A_t - F_{jth}) \tag{16}$$

Barry and Jennings (1992) derived a similar relationship among uncertainty, disagreement and the average covariance in forecasts. Their result justifies forecast disagreement as one component of forecast uncertainty, which has, unfortunately, been unnoticed in the economics literature. Given our model assumptions, we can simplify the expression for the average covariance among forecast errors in (16) as

$$C_{th} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N E[(\lambda_{th} + \varepsilon_{ith})(\lambda_{th} + \varepsilon_{jth})] = \sigma_{\lambda|th}^2 \tag{17}$$

which can be easily interpreted as the uncertainty shared by all forecasters due to their exposure to common shocks. Thus, (17) greatly simplifies the results in Barry and Jennings (1992) and Barron *et al.* (1998), and gives the relationship (8).

<sup>6</sup> Our measure of uncertainty is different from the ‘combined uncertainty’ as defined by the variance of aggregate density forecast in Wallis (2005), which includes both our measure of uncertainty and the disagreement as its components.

### 2.3. Relationship between Uncertainty and Disagreement in Real Time

Note that equation (8) specifies a relationship between uncertainty, disagreement and the variance of aggregate shocks based on unconditional expectations before observing any forecast or actual value. However, the forecasts and the associated uncertainty are generated sequentially in the real time. Thus, we should develop a corresponding relationship in terms of expectations conditional on observing the individual forecasts (and hence disagreement  $d_{th}$ ) at time  $t - h$ , but before the actual value  $A_t$  was realized.

Following Engle (1983), we decompose the average squared individual forecast errors as

$$\frac{1}{N} \sum_{i=1}^N (A_t - F_{ith})^2 = (A_t - F_{.th})^2 + \left(1 - \frac{1}{N}\right) d_{th} \quad (18)$$

Taking expectations on both sides given all available information at time  $t$  including  $F_{ith}$  and  $d_{th}$ , we obtain the following conditional relationship between aggregate uncertainty, the variance of consensus forecast errors and observed disagreement:

$$U_{th} = E(A_t - F_{.th})^2 + \left(1 - \frac{1}{N}\right) d_{th} \quad (19)$$

The first term on the right-hand side of (19) can alternatively be written as (see Markowitz, 1959, p. 111)<sup>7</sup>

$$E(A_t - F_{.th})^2 = \frac{1}{N^2} E \left[ \sum_{i=1}^N (A_t - F_{ith})^2 \right] + \frac{1}{N^2} E \left[ \sum_{i=1}^N \sum_{j \neq i}^N (A_t - F_{ith})(A_t - F_{jth}) \right] \quad (20)$$

Given our framework, (20) can be expressed as

$$E(A_t - F_{.th})^2 = \sigma_{\lambda|th}^2 + \frac{1}{N^2} \sum_{i=1}^N \sigma_{\varepsilon|ith}^2 \quad (21)$$

We should point out that the uncertainty about the consensus forecast in (21) defined by Bomberger (1996) is different from our measure of forecast uncertainty in (7). The uncertainty about the consensus forecast is less than the average of the individual uncertainties due to the fact that combining individual forecasts implicitly pools the diverse idiosyncratic errors. Note that, as the number of forecasters goes to infinity, the uncertainty about the consensus forecast will reflect only the uncertainty in the common information.

Substituting (21) in (19), we obtain

$$U_{th} = \sigma_{\lambda|th}^2 + \frac{1}{N^2} \sum_{i=1}^N \sigma_{\varepsilon|ith}^2 + \left(1 - \frac{1}{N}\right) d_{th} \quad (22)$$

<sup>7</sup> In the context of forecast combination, Batchelor and Dua (1995) had a similar decomposition.

For typical values of  $N$  and  $\sigma_{\varepsilon|ith}^2$  in our context, the second term on the right-hand side of (22) will be very close to zero and can be ignored.<sup>8</sup> Thus the difference between the reported *ex ante* forecast uncertainty and disagreement will give approximate estimates of *ex ante* variances of aggregate shocks in real time before the actual values are realized.

In the context of equations (21) and (22), one can understand the efforts of Bomberger (1996), who examined the dependence of uncertainty associated with the average forecast (which he called ‘consensus uncertainty’) on forecast disagreement. Certainly, a positive relationship between the two during periods of economic instability will ensure that disagreement will continue to be positively correlated with the overall forecast uncertainty. However, since the difference between uncertainty and disagreement is the variance of unanticipated aggregate shocks, theoretically it is not clear why disagreement will be able to predict the latter. Our model assumptions, though admittedly simple, rule out any feedback from perceived future variability of common shocks to current idiosyncratic individual variances. However, it is possible that greater uncertainty about future common shocks affects current individual  $\sigma_{\varepsilon|ith}^2$  and hence the latter covaries with  $\sigma_{\lambda|th}^2$ . This is how Bomberger’s (1996) econometric exercise can be justified.

On the other hand, as Zarnowitz and Lambros (1987) have pointed out, there may be periods where all forecasters agree on relatively high macroeconomic uncertainty in the immediate future, and hence disagreement between forecasters will be low even though uncertainty is high. The opposite is also possible, where forecasters disagree a lot about their mean forecasts but are confident about their individual predictions. This situation will arise when forecasters disagree on otherwise precise models and scenarios that should be used to depict the movement of the economy over the forecasting horizon. Thus, lacking any theoretical basis, the strength and stability of the relationship between disagreement and overall forecast uncertainty becomes an empirical issue. But our result clearly suggests that the relationship will depend crucially on the sample period, the target variable, and length of the forecast horizon. Our analysis also helps to reconcile the divergent findings in previous empirical studies examining the appropriateness of disagreement as a proxy for forecast uncertainty. Certainly, contrary to a statement in Bomberger (1996, p. 385), it is *not* necessary that ‘if disagreement is to be a good proxy for individual uncertainty, it must also track consensus uncertainty’.

### 3. EMPIRICAL TESTS OF THE RELATIONSHIP BETWEEN UNCERTAINTY AND DISAGREEMENT

This section begins with a short description of data on density forecasts used in this study. In subsequent sections, we present empirical evidence in support of our hypothesized relationship between disagreement and uncertainty over time and horizons. We then evaluate the appropriateness of using squared error of mean forecasts as a proxy for the variance of aggregate shocks that has been extensively used in recent accounting literature. Our suggestion to construct a robust measure of *ex ante* uncertainty, given a panel of forecasts, is presented at the end.

---

<sup>8</sup> In our sample, the average values of  $\frac{1}{N^2} \sum_{i=1}^N \sigma_{\varepsilon|ith}^2$  lie between 0.01 and 0.02 for both inflation and output growth forecasts.



### 3.1. Data

The data in our study are taken from the Survey of Professional Forecasters (SPF), provided by the Federal Reserve Bank of Philadelphia. A unique feature of SPF data is that forecasters are also asked to provide density forecasts for output growth and inflation, which is the focus of this paper. The historical time series of forecasts in this survey is quite lengthy (since the fourth quarter of 1968), and there are a number of changes in the surveys that make the data challenging to work with. We focus on the density forecasts for the change from year  $t - 1$  to  $t$  that were issued in the four consecutive surveys from the first quarter through the fourth quarter of year  $t$ . The actual horizons for these four forecasts are approximately 3 1/2, 2 1/2, 1 1/2, and 1/2 quarters but we shall refer to them simply as horizons 4, 3, 2, and 1 quarter. After deleting observations with missing values, we obtain a total of 4986 observations for inflation over 1969:Q1 to 2007:Q4 and 3312 observations for output growth over 1981:Q3 to 2007:Q4.<sup>9</sup> For the purpose of estimation, we eliminate observations of infrequent respondents. We focus on the ‘regular’ respondents who participated in at least 25 surveys in inflation forecasts and at least 17 surveys in output growth forecasts—approximately 15% in both cases. This leaves us with a total of 2787 observations for inflation forecasts and 2342 observations for output growth forecasts.<sup>10</sup>

To test the hypothesized relationships, we also need the actual values of inflation and output growth. As is well known, the NIPA data often go through serious revisions. Obviously, the most recent revision is not a good choice, since it involves adjustment of definitions and classifications. Consistent with the findings in Harvey and Newbold (2003) that the unrevised data approximates the forecasters’ objective better, we choose the first release of annual inflation and output growth to compute the actual values. These are the real-time data available from the Federal Reserve Bank of Philadelphia.<sup>11</sup>

### 3.2. Estimates of Uncertainty and Disagreement

Note that the variance of forecast error in (6) can be interpreted as the variance of random variable  $A_t$  as perceived by individual  $i$ , given information available at time  $t - h$ , which is conceptually the same as the variance of the density forecast reported by individual  $i$ . Taking the average of the variances of individual densities yields estimates of forecast uncertainty as defined in (7).

To obtain appropriate measures of forecast disagreement, we need to control for possible individual biases in the sample. These biases may arise due to different loss functions held by forecasters and other behavioral factors that are expected to be relatively stable over time. Following Davies and Lahiri (1995, 1999), the individual forecast error has a three-dimensional nested structure in the presence of individual bias  $\phi_{ih}$ :

$$e_{ith} \equiv A_t - F_{ith} = \phi_{ih} + \lambda_{ih} + \varepsilon_{ith} \quad (23)$$

<sup>9</sup> The Philadelphia Fed is uncertain about the target years referred to in the surveys made in the first quarter of 1985 and 1986. We deleted those forecasters who were obviously misled by the wrong wording of the question and used the rest of the responses.

<sup>10</sup> See Giordani and Söderlind (2003) and Lahiri and Liu (2005) for a detailed discussion on the specification and construction of the analytical sample.

<sup>11</sup> All calculations reported in this paper were also repeated with the so-called ‘first final’ (i.e., the third monthly revision) and July revisions. The main results and conclusions were unchanged.

The systematic individual bias,  $\hat{\phi}_{it}$ , can be estimated as

$$\hat{\phi}_{it} = \frac{1}{T} \sum_{t=1}^T (A_t - F_{it}) \quad (24)$$

After adjusting for these individual biases, we obtain forecast disagreement based on unbiased forecasts.<sup>12</sup>

Estimates of uncertainty, disagreement and their difference, which is an estimate of the variance of *ex ante* aggregate shock, are plotted in Figures 1–4. Their average values are given in Table I. Several points are worth noting. Disagreement and uncertainty typically move together but the former is almost always smaller than the latter in both series, which is in line with the evidence that the former tends to underestimate the latter (cf. Zarnowitz and Lambros, 1987; Lahiri *et al.*, 1988). Uncertainty is seen to be very sticky in terms of its high autocorrelation and low volatility and, as a result, responds slowly to even rapid changes in the economic environment.<sup>13</sup> Also, the difference between uncertainty and disagreement (i.e., the variance of *ex ante* aggregate shocks) in both series becomes larger, as forecast horizon gets longer from one quarter to four quarters, which provides evidence in support of the horizon effect. Note also that the estimated variances of aggregate shocks are systematically much bigger for GDP growth than inflation at all horizons. This finding implies that it is more difficult to forecast real GDP growth than inflation, and is consistent with most studies on forecast evaluation that report significantly higher RMSE for real GDP than for inflation forecasts.<sup>14</sup>

Second, Figures 1–4 suggest that the volatility of aggregate shocks declined sharply after 1991 for both inflation and output growth. This finding contributes to our understanding of the factors behind the *Great Moderation*—the well-documented decline in macroeconomic volatility in the USA since 1984. Our result suggests that the decline in macroeconomic volatility during 1984–1991 cannot be attributed to ‘good luck’, since the economy was hit by unforeseen large shocks during this period (cf. Campbell, 2007), and instead must be explained by other factors, such as structural changes (cf. McConnell and Perez-Quiros, 2000) or improved monetary policy (cf. Mishkin, 2007). After 1991, the shocks hitting the US economy have become smaller and more stable, and have thus played a large role in the reduction of macroeconomic volatility.

Third, somewhat unexpectedly, in some quarters disagreement exceeds uncertainty, especially for inflation. Certainly, one reason can be the imprecision in the estimation of uncertainty and disagreement based on a finite sample of survey respondents. After all, relationships (8) and (22) are expected to hold only on the average. It is well known that the distribution and variance of point forecasts can be particularly volatile due to the presence of few outliers in the sample. Indeed, if we use more robust measures of disagreement such as the quasi-standard deviation or inter-quartile

<sup>12</sup> In the presence of individual bias ( $\phi_{it}$ ), disagreement becomes  $D_{it} = \sigma_{\phi}^2 + \sum_{i=1}^N \sigma_{\varepsilon_{it}}^2 / N$ , where  $\sigma_{\phi}^2$  is the sample variance of individual bias across forecasters. Individual biases were, however, estimated to be relatively small, and thus they did not affect forecast disagreement by any significant amount.

<sup>13</sup> This is also true for time series measures of uncertainty. Giordani and Söderlind (2003) and Lahiri and Liu (2006) show that the GARCH measure of uncertainty fails to capture the increase in inflation uncertainty around the second oil price shock.

<sup>14</sup> See, for instance, Öller and Barot (2000), Banerjee and Marcellino (2006), Reifschneider and Tulip (2007), and Lahiri and Sheng (2010).

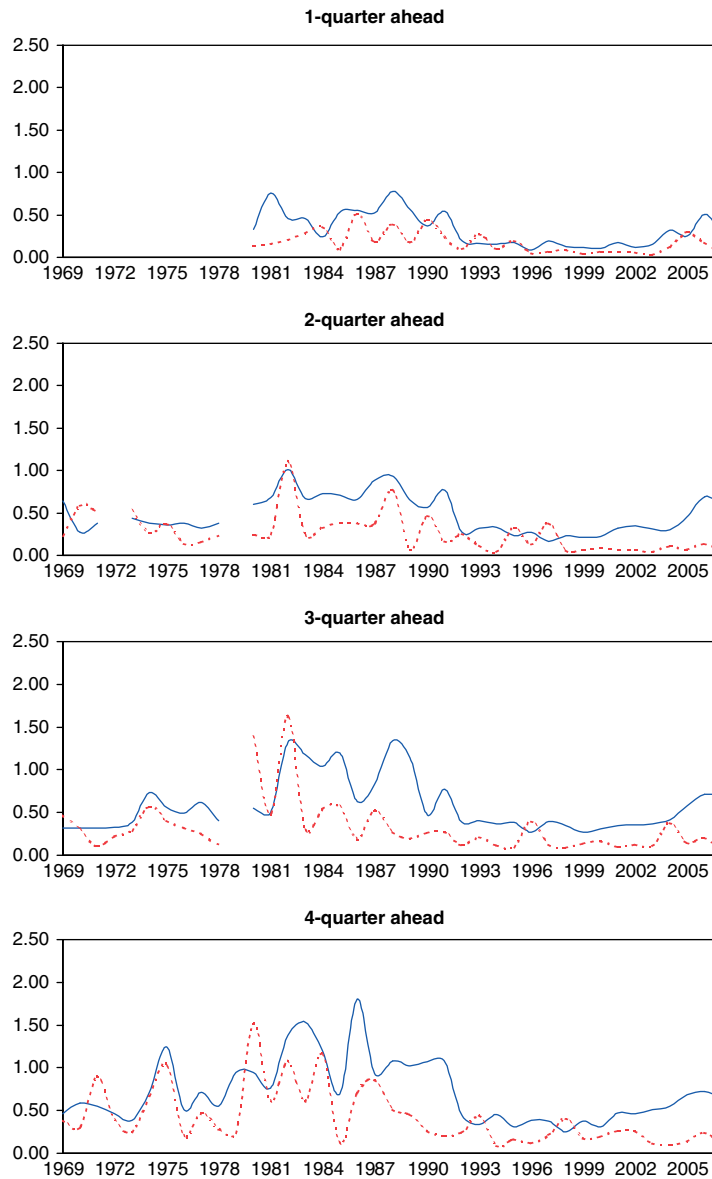


Figure 1. Uncertainty (solid line) and disagreement (dotted line) in inflation forecasts. This figure is available in color online at [www.interscience.wiley.com/journal/jae](http://www.interscience.wiley.com/journal/jae)

range, this apparent anomaly mostly disappears (see Figure 5 in Giordani and Söderlind (2003) paper).<sup>15</sup> There are other possibilities that should also be pointed out. It could be that survey

<sup>15</sup> Quasi standard deviation is the half distance between the 84th and 16th percentiles of the distribution of point forecasts. Under normality, it will be identical to the usual standard deviation.

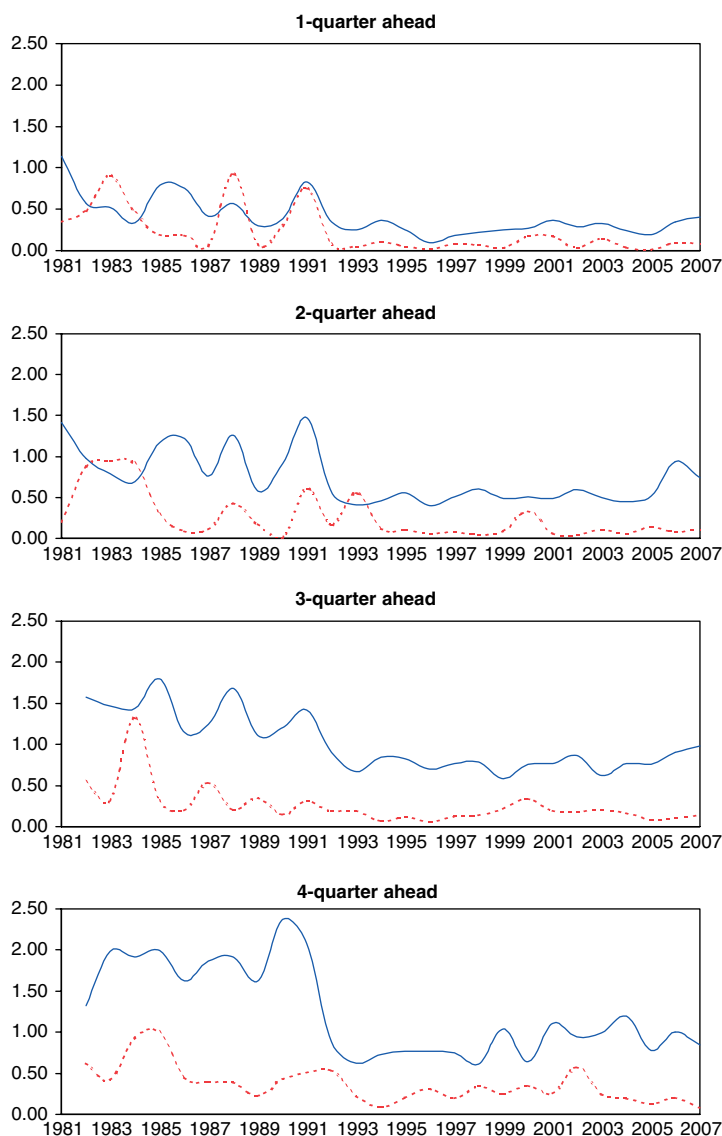


Figure 2. Uncertainty (solid line) and disagreement (dotted line) in real GDP forecasts. This figure is available in color online at [www.interscience.wiley.com/journal/jae](http://www.interscience.wiley.com/journal/jae)

measure of uncertainty does not represent the ‘true’ or objective uncertainty correctly. Diebold *et al.* (1999) concluded that survey uncertainty overestimates the true values. However, Giordani and Söderlind (2003) reached an opposite conclusion. Following the latter approach, in Table II we report the average percentage times the 90% predictive interval covers the actual outcomes after fitting a uniform distribution over the bins during 1969–2007. We find that survey measures of uncertainty are well calibrated for all horizons except four-quarter-ahead forecasts, for which

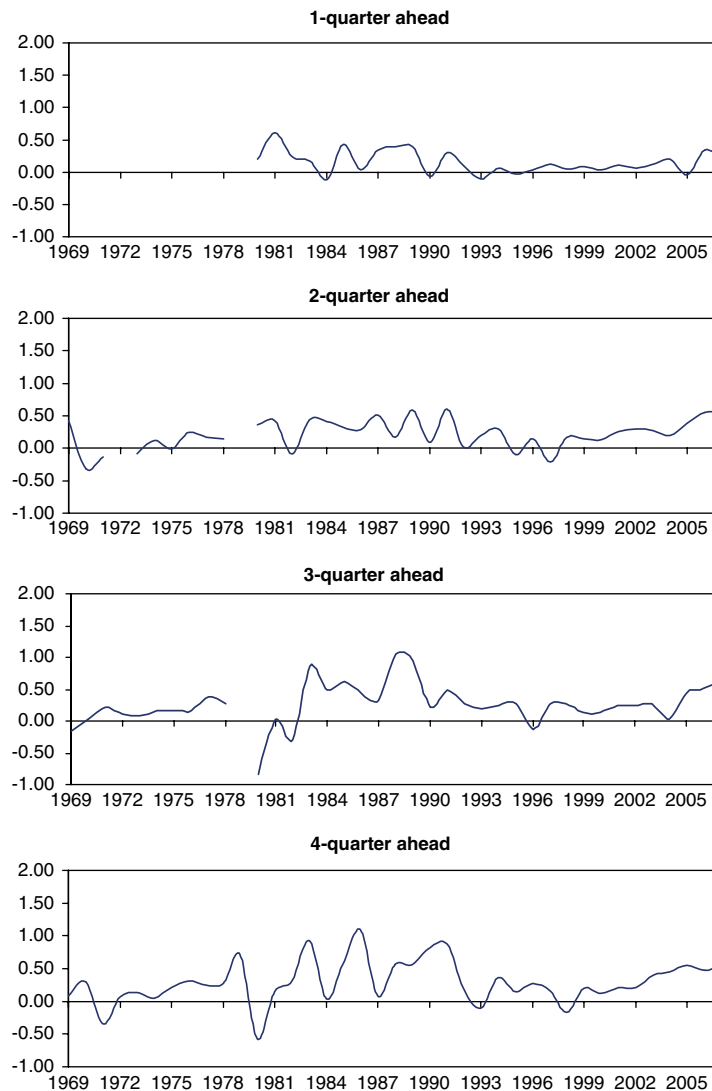


Figure 3. Estimated variance of aggregate shocks in inflation forecasts. This figure is available in color online at [www.interscience.wiley.com/journal/jae](http://www.interscience.wiley.com/journal/jae)

the survey measure underestimates the objective uncertainty by 13% for inflation and 17% for output growth forecasts. This possible underestimation of the true uncertainty by survey densities can rule out a few of the negative estimates of the variance of aggregate shocks.<sup>16</sup> Also, if we believe that, for a particular horizon, the extent of under- or over-estimation is time invariant, the

<sup>16</sup> Following Giordani and Söderlind (2003), we also fitted normal distributions over histograms and repeated the same comparison exercise. As expected, the normal approximation suggested even more underestimation. Many recent studies have, however, avoided the practice of fitting normal distribution to the individual density forecasts because the majority of the respondents seldom assign probabilities to more than three intervals (see Engelberg *et al.*, 2009).

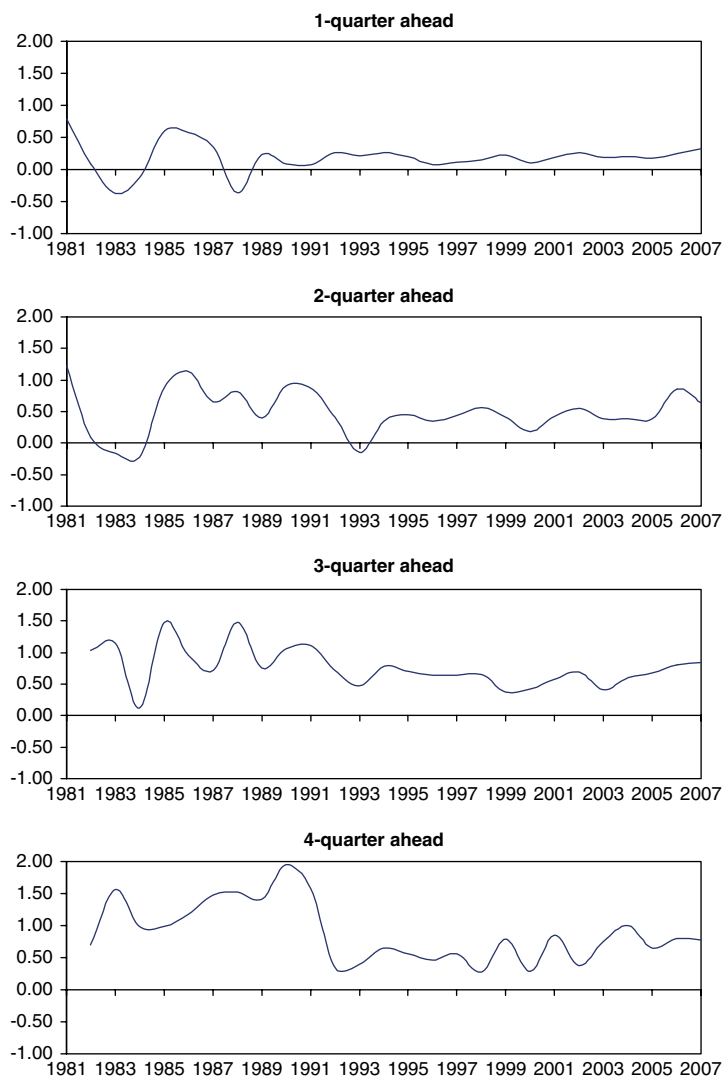


Figure 4. Estimated variance of aggregate shocks in output growth forecasts. This figure is available in color online at [www.interscience.wiley.com/journal/jae](http://www.interscience.wiley.com/journal/jae)

survey uncertainty will continue to be a meaningful indicator for true forecast uncertainty. Even if adjusted upwards by 13%, the inflation uncertainty is still far less than disagreement for the four-quarter-ahead forecast for 1980.

Another explanation can be the well-documented structural break of the early 1980s. As is well known, inflation rose sharply and unexpectedly during 1979–1981, and was characterized by a break in the inflation process. This may lead to forecasters adopting disparate forecasting functions and, as a result, their predictions may generate extraordinary disagreement. As Mankiw *et al.* (2003) have argued, the distribution of point forecasts may become multimodal as forecasters

Table I. Uncertainty and disagreement averaged over time

	SPF inflation forecast (1969–2007)				SPF GDP forecast (1981–2007)			
	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Uncertainty	0.33	0.48	0.58	0.69	0.41	0.74	1.02	1.25
Disagreement	0.18	0.26	0.32	0.42	0.22	0.25	0.26	0.37
Difference	0.15	0.22	0.26	0.27	0.19	0.49	0.76	0.88

learn about the new regime and react differently. More specifically, this may lead to significant correlations—some negative and some positive—in idiosyncratic shocks between the forecasters, violating one of the assumptions in our forecast error decomposition, viz.,  $E(\varepsilon_{ith}\varepsilon_{jth}) = 0$  for any  $t$ ,  $h$  and  $i \neq j$ . Note that, as equations (5) and (16) reveal, for the validity of relation (8) or (22), we need the average pair-wise covariances in forecast errors (net of common shocks) to be zero. Using the four-quarter-ahead inflation forecasts, however, we found that this condition was not violated even during the tumultuous period 1979–1983.<sup>17</sup> Thus, on balance, we feel that the estimated negative variance of aggregate shocks on few occasions in our sample can be attributed to the imprecision in the estimation of disagreement or uncertainty due to small samples and outliers.

We now formally test the implications of (22) that the relationship between uncertainty and disagreement depends on the variance of aggregate shocks over time and across horizons. By plotting the actual inflation rate, we find its average value during 1969–1983 to be at least 2.5 times that during 1984–2007, consistent with the stylized fact documented in the literature (cf. Stock and Watson, 2007). As is well known, higher rates of inflation are generally associated with higher variability of inflation and presumably greater uncertainty about future rates. We thus divide the sample of inflation forecasts into two periods: the unstable period (1969–1983) and the stable period (1984–2007). To study the relationship between uncertainty and disagreement, we run the following regression:

$$U_{th} = \beta D_{th} + \rho_1 H_1 + \rho_2 H_2 + \rho_3 H_3 + \rho_4 H_4 + \varepsilon_{th} \quad (25)$$

where  $H_i = 1$  if the forecast is made at horizon  $i$  for  $i = 1, 2, 3, 4$ , and 0 otherwise.

Table III shows the estimation results. The estimated coefficient on disagreement is 0.43 for inflation forecasts during 1969–1983. The same coefficient during 1984–2007 is estimated to be 0.76 and 0.72 for inflation and GDP forecasts, respectively. Thus the evidence from SPF density forecasts supports our model implication that disagreement is a good proxy for uncertainty when the variance of aggregate shocks is small, and is consistent with the empirical results presented by Bomberger (1996) and Giordani and Söderlind (2003). As is also clear in Table III, the difference between uncertainty and disagreement, which is an estimate of *ex ante* variance of aggregate shocks, is larger, as forecast horizon gets longer. For example, as the horizon increases from one quarter to four quarters, the difference increases monotonically from 0.24 to 0.96 in output growth forecasts. This pattern is also observed for inflation forecasts during the stable period at all horizons, with the exception of four-quarter-ahead forecasts,

<sup>17</sup> Even after requiring a minimum of four common observations to calculate each pair-wise correlation, we had only seven forecasters during 1976–1983. We also calculated the average pair-wise correlations during a relatively stable period (1993–2000, which excluded two bordering recessions) with 16 available forecasters. The average values of the covariances were  $-0.04$  and  $-0.06$ , respectively, during these two fairly diverse periods.

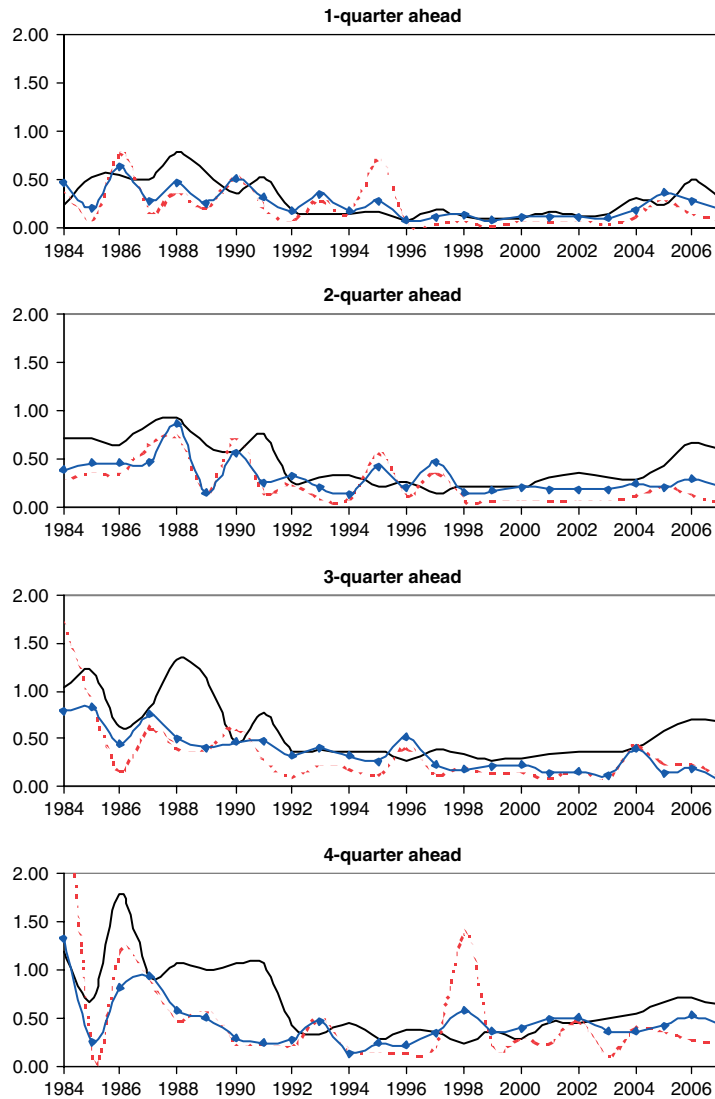


Figure 5. Measures of uncertainty in inflation forecasts: survey measure of uncertainty (solid line); uncertainty using squared error of mean forecast (dotted line); uncertainty from GARCH-type model (line with diamond). This figure is available in color online at [www.interscience.wiley.com/journal/jae](http://www.interscience.wiley.com/journal/jae)

which means that the additional variability due to the shocks that fell during the first quarter of the current year (on average during 1984–2007) compared to the remaining quarters is not significant. This is caused by the relatively high disagreement in four-quarter-ahead forecasts during the 1986–1989 period compared to other forecasts (see Figure 1). Furthermore, all horizon dummies are estimated to be statistically significant at the 5% level. On balance, the empirical evidence above shows that the variance of aggregate shocks accumulates systematically over horizons, as predicted by our model. This finding is important since most of studies



Table II. Comparison of 90% predictive interval with actual outcomes

Horizon	4Q ahead	3Q ahead	2Q ahead	1Q ahead
SPF inflation	78.48	85.46	89.34	88.22
SPF output growth	74.35	87.17	85.95	83.33

*Note:* This table shows the percentage of times that the 90% predictive interval covers the actual outcomes. Predictive intervals are constructed from SPF individual density forecasts during 1969–2007 by fitting uniform distribution over histograms.

Table III. Regression of survey measure of uncertainty on disagreement over time

	SPF inflation forecast		SPF GDP forecast
	1969–1983	1984–2007	1984–2007
Disagreement	0.43* (0.08)	0.76* (0.19)	0.72* (0.25)
H1	0.39* (0.13)	0.17* (0.03)	0.24* (0.04)
H2	0.34* (0.04)	0.31* (0.04)	0.56* (0.05)
H3	0.36* (0.06)	0.42* (0.04)	0.81* (0.06)
H4	0.53* (0.06)	0.39* (0.07)	0.96* (0.09)
Adj. $R^2$	0.34	0.39	0.53

*Note:* Standard errors are in parentheses.

\* Estimated values are significant at the 5% critical level.

have focused on their relationship over time, without specifying the underlying forecast horizons.<sup>18</sup>

### 3.3. Should Squared Error of Mean Forecast be Used as a Proxy for $\sigma_\lambda^2$ ?

An influential paper in the accounting literature by Barron *et al.* (1998) extended the model in Barry and Jennings (1992) and suggested that ‘one can infer uncertainty and consensus from observable forecast dispersion, error in the mean forecast and the number of forecasts’ (Barron *et al.*, 1998, p. 427). Their suggestion has been extensively used to study the information environment in analysts’ earning forecasts. Yet, without direct information on uncertainty, the validity of their suggestion in finite samples can never be established. Our analysis below addresses this issue.

Barron *et al.* (1998) argued that one could use the squared error in the mean forecast as a proxy for the variance of aggregate shock to empirically estimate forecast uncertainty as in the following equation:

$$\hat{U}_{it} = (A_t - F_{.it})^2 + \left(1 + \frac{1}{N}\right)d_{it} \quad (26)$$

<sup>18</sup> Two exceptions are the recent papers by Patton and Timmermann (2008) and Lahiri and Sheng (2008), who explicitly modeled the term structure of survey forecasts over horizons.

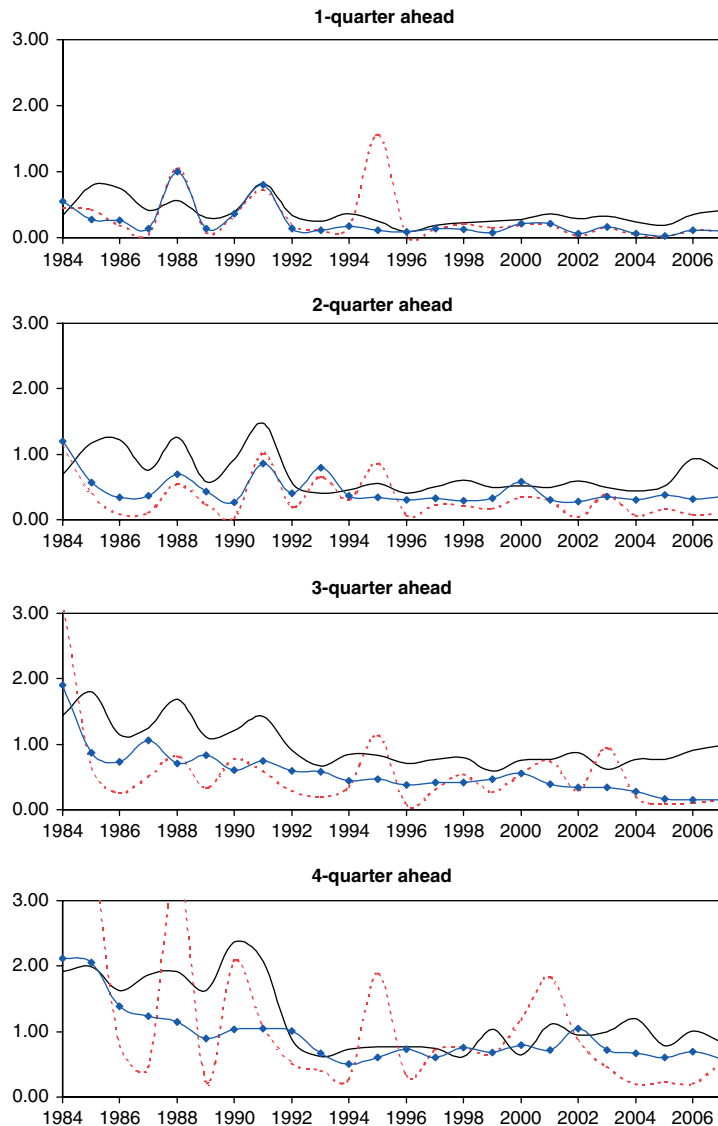


Figure 6. Measures of uncertainty in output growth forecasts: survey measure of uncertainty (solid line); uncertainty using squared error of mean forecast (dotted line); uncertainty from GARCH-type model (line with diamond). This figure is available in color online at [www.interscience.wiley.com/journal/jae](http://www.interscience.wiley.com/journal/jae)

Comparing the above equation to (18), it immediately follows that the measure of uncertainty in (26) is nothing but the average squared individual forecast errors.<sup>19</sup> Because forecast errors are

<sup>19</sup> During our sample period, the squared error of the mean forecast accounts for 40–70% of *ex post* uncertainty in output growth forecast and for 30–60% in inflation forecast, as the horizon gets longer from one to four quarters ahead. The remainder is attributable to disagreement.

Table IV. Regression of survey measure of uncertainty on *ex post* uncertainty

	SPF inflation forecast		SPF GDP forecast
	1969–1983	1984–2007	1984–2007
<i>Ex post</i> uncertainty	0.02 (0.02)	0.27* (0.07)	0.25* (0.05)
H1	0.46* (0.22)	0.24* (0.02)	0.30* (0.01)
H2	0.49* (0.05)	0.40* (0.02)	0.62* (0.02)
H3	0.55* (0.06)	0.49* (0.03)	0.84* (0.03)
H4	0.75* (0.06)	0.48* (0.05)	0.93* (0.06)
Adj. $R^2$	0.09	0.30	0.54

Note: Standard errors are in parentheses.

\* Estimated values are significant at the 5% critical level.

known to respondents only after the announcement of actual values, (26) indeed yields a measure of *ex post* uncertainty. Its reliability as a proxy for *ex ante* uncertainty faced by individual forecasters at the time of forecast is questionable. With density forecasts at our disposal, we can compare them directly.

Figures 5 and 6 plot these two measures of uncertainty during 1984–2007 for inflation and output growth forecasts, respectively. The general message is that, compared to survey measure of uncertainty, *ex post* uncertainty based on (26) is considerably more volatile. In particular, the *ex post* uncertainty overstates the survey measure of uncertainty whenever a forecast is followed by a large unanticipated forecast error. This is unfortunate because, being unanticipated, these errors should not have affected the forecast uncertainty that pre-dates the observed forecast error. The regression results in Table IV reinforce some of the features from these graphs. For inflation forecasts, the estimated coefficient on *ex post* uncertainty is almost zero during the unstable period 1969–1983. Even in the stable period, the coefficients are estimated to be very small for both inflation and output growth forecasts. Comparing  $\bar{R}^2$  in Tables III and IV, we see that disagreement alone is a reasonable proxy for uncertainty. However, adding the squared error in the consensus forecast to disagreement turns out to be a significantly worse proxy for uncertainty than the disagreement alone.  $\bar{R}^2$  falls from 0.34 to 0.09 during 1969–1983 and from 0.39 to 0.30 during 1984–2007 for inflation, implying that the squared forecast errors contribute negatively to explaining survey uncertainty. For real GDP, the squared forecast errors have practically no additional explanatory power, as  $\bar{R}^2$  increases from 0.53 to 0.54.

Clearly, forecast uncertainty constructed according to Barron *et al.* (1998) depends on the realization of individual forecast errors. But forecast error is necessarily an *ex post* quantity, which reflects unexpected shocks after the forecast is made, and thus should not affect uncertainty at the time a forecast is issued. One may think that it may be an acceptable practice to use mean squared forecast error as a proxy for its *ex ante* counterpart because Barron *et al.* (1998) are looking at forecast uncertainty retrospectively. Their measure has been used to study the impact of special events, such as Regulation Fair Disclosure, on the forecasting environment of financial analysts (see, for example, Mohanram and Sunder, 2006, and references therein). Even in this

retrospective context, the squared forecast error can give a very misleading indication of the uncertainty environment in real time in a past sample, as shown by the extra variability in *ex post* uncertainty during periods that are characterized by large *ex post* forecast errors (see Figures 5 and 6).

Engle (1983) demonstrated that the average squared individual forecast errors do not show patterns similar to ARCH measures of uncertainty.<sup>20</sup> Our findings here, together with the empirical evidence presented in Engle (1983), strongly caution against using the squared error in the mean forecast as a component of overall forecast uncertainty. We show below that forecast disagreement by itself, without adding the *ex post* mean squared error, correlates better with the observed survey uncertainty.

### 3.4. *Ex Ante* Measures of Uncertainty using GARCH

Because uncertainty is essentially an *ex ante* concept attached to a forecast before the actual outcome is known, it must be constructed using data available in real time. The celebrated ARCH model (Engle, 1982) and its various extensions are now standard methods for modeling forecast uncertainty based on aggregate data. This literature posits that forecast uncertainty of a variable can be measured by the conditional variance of its forecast error that, in turn, is assumed to depend on past forecast error and lagged forecast uncertainty.

In order to generate GARCH-type estimates of the variability of aggregate shocks, we first filter the mean forecast errors for possible autocorrelation; see Harvey and Newbold (2003). The order of autocorrelation present in a given mean forecast error series is found by fitting moving average models of varying order, the preferred model being chosen by the use of Schwarz information criterion. We then estimated  $\sigma_{\lambda|t}^2$  using different GARCH-type models with various distributional assumptions on the filtered mean forecast errors. For convenience, these models are labeled as Model 1 through Model 8. In Model 1, we estimated the standard GARCH (1, 1) model with the following specification:

$$e_t \sim N(0, \sigma_{\lambda|t}^2), \quad \sigma_{\lambda|t}^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 \sigma_{\lambda|t-1}^2 \quad (27)$$

where  $e_t$  is the serially uncorrelated mean forecast error. Equation (27) has been estimated using the quasi-maximum likelihood procedure (cf. Bollerslev and Wooldridge, 1992) for the 1984–2007 subsample and for each horizon. Consistent with many earlier studies, in Model 2 we estimated (27) using the  $t$ -distribution with six degrees of freedom. As an alternative specification,<sup>21</sup> we replaced the lagged mean squared forecast error in (27) with the average of mean squared errors over the

<sup>20</sup> As shown in Table II of Engle (1983), the average squared individual forecast errors are 31.78 (1947/12–1952/6), 1.35 (1962/6–1966/12) and 13.01 (1971/6–1975/12), but the corresponding ARCH uncertainty is 19.22, 2.57 and 3.37, respectively.

<sup>21</sup> This follows Reifschneider and Tulip (2007), who suggested a measure of past forecast uncertainty using squared individual forecast errors of a number of private and government forecasters averaged over 1986–2006. Their purpose is to use this average historical uncertainty based on past predictive accuracy as a benchmark against which FOMC participants can assess their present uncertainty.

last 10 years.<sup>22</sup> Thus, in Model 3, we estimated  $\sigma_{\lambda|t}^2$  using the following model specification:

$$e_t \sim N(0, \sigma_{\lambda|t}^2), \quad \sigma_{\lambda|t}^2 = \beta_0 + \beta_1 \left( \sum_{s=1}^{10} MSE_{t-s} / 10 \right) + \beta_2 \sigma_{\lambda|t-1}^2 \quad (28)$$

Model 4 estimated (28) using the  $t$ -distribution with six degrees of freedom. Models 5–8 correspond to Models 1–4, except that we modeled the standard deviation instead of the variance in the GARCH-type models. The estimation results, not reported here, show that the lagged variance of aggregate shocks was significant at the 5% level in the majority of cases, but the lagged mean forecast errors, as well as the average of mean squared errors over the last 10 years, are only significant in some cases, depending on the horizons and variables under study.<sup>23</sup>

To form a measure of forecast uncertainty, our suggestion is to use the observed disagreement from the survey  $d_{th}$  and the variance of aggregate shocks generated conditionally by GARCH-type models  $\hat{\sigma}_{\lambda|th}^2$  to estimate  $U_{th}$ :

$$\hat{U}_{th} = \hat{\sigma}_{\lambda|th}^2 + \left(1 - \frac{1}{N}\right) d_{th} \quad (29)$$

The justification is as follows. Uncertainty comes from two sources: the error components in common information and in private information. The  $\hat{\sigma}_{\lambda|th}^2$  term captures the imprecision in common information, and  $d_{th}$  reflects the imprecision in forecasters' idiosyncratic information and diversity in forecasting models. The measure of uncertainty in (29) avoids the drawback of the inability to capture the heterogeneity of forecasting models in using the GARCH measure of uncertainty alone. Our suggestion is supported by the findings in Batchelor and Dua (1993) and Bomberger (1996); in a comparison of ARCH and survey measures of uncertainty, these two studies concluded that the former tends to be lower than the latter and, more importantly, the former is less variable over time than the latter. Thus, if one accepts survey measures as valid, ARCH measure alone underestimates the level and the variation in uncertainty over time.

According to (29), forecast uncertainty is generated by the sum of the estimated variance of aggregate shocks from GARCH-type models and the disagreement from the survey. Table V shows the correlations between survey and generated measures of uncertainty. Several points stand out. First, the GARCH estimates of uncertainty with the average squared errors over the last 10 years (in place of the last period forecast error) help to capture the variation in the survey measure of uncertainty fairly well (Models 3, 4, 7 and 8). Compared to the simple correlation with the disagreement alone (the first row in Table V), the correlations between the survey uncertainty and the uncertainty generated by Models 3, 4, 7 and 8 increase by about 5% for one- and two-quarter-ahead inflation forecasts, and by more than 15% and 10% for three- and four-quarter-ahead GDP forecasts, respectively. This is because the relative importance of aggregate shocks is more for GDP growth than inflation forecasts, and hence it pays to project them using GARCH. Second, models with  $t$ -distributions (Models 2, 4, 6 and 8) match survey measure of uncertainty better. In general, Models 2, 4, 6 and 8 using  $t$ -distribution with six degrees of freedom perform better in capturing the variation in survey uncertainty than Models 1, 3, 5 and 7 using normal distribution. Third,

<sup>22</sup> During 1974–1981, SPF did not ask for the annual average forecast. We matched the reported quarterly point forecasts with the real-time data to derive the implied annual forecasts for the current year.

<sup>23</sup> Following Bomberger (1996), we also added disagreement in the variance equation of the GARCH models and found that disagreement never became significant at the 5% level. This is consistent with the findings in Rich and Butler (1998).

modeling the standard deviation instead of the variance tends to do a better job in representing the variation in the survey measure of uncertainty. For output growth forecasts, the best model seems to be Model 8, which performs even better at longer horizons. For inflation forecasts, the best model is Model 8 at shorter horizons and Model 6 at longer horizons. In addition, when we add squared mean forecast errors to disagreement (Model 0), its predictive power to proxy survey uncertainty decreases across almost all horizons for both variables, echoing a point that we have established in Section 3.3.

In summary, compared to squared forecast errors, the GARCH-type models are relatively more successful in modeling the variability of future aggregate shocks to the economy in the sense that, when added to disagreement, this composite measure of *ex ante* forecast uncertainty explains the corresponding survey measure better than disagreement alone. However, the additional explanatory power due to the addition of GARCH estimates is sometimes modest.<sup>24</sup>

We plot the evolution of uncertainty generated from the best models in inflation and output growth forecasts over time in Figures 5 and 6. Compared to the uncertainty constructed using the

Table V. Correlation between survey uncertainty and alternative measures of uncertainty

	SPF inflation forecast (1984–2007)				SPF GDP forecast (1984–2007)			
	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Disagreement	0.56	0.52	0.55	0.60	0.60	0.30	0.44	0.58
Model 0	0.36	0.49	0.56	0.52	0.32	0.24	0.39	0.57
Model 1	0.56	0.51	0.67	0.44	0.63	0.33	0.46	0.42
Model 2	0.59	0.51	0.67	0.53	0.62	0.32	0.42	0.47
Model 3	0.62	0.54	0.64	0.51	0.56	0.24	0.60	0.70
Model 4	0.61	0.54	0.61	0.53	0.61	0.31	0.62	0.71
Model 5	0.57	0.52	0.66	0.49	0.58	0.20	0.37	0.50
Model 6	0.57	0.53	0.67	0.54	0.63	0.34	0.39	0.33
Model 7	0.62	0.56	0.61	0.52	0.58	0.37	0.62	0.69
Model 8	0.63	0.56	0.61	0.51	0.61	0.34	0.61	0.71

*Note:* This table presents the correlations between survey and alternative measures of uncertainty. Alternative measures of uncertainty are generated by the sum of the variance of aggregate shocks from Models 0 to 8 and the disagreement from the survey. In particular, in Model 0, the squared error in the mean forecasts is used as a proxy for the variance of aggregate shocks. In Models 1–8, the variance of aggregate shocks is generated from the following models:

Model 1: GARCH (1, 1) with normal distribution.

Model 2: GARCH (1, 1) with *t*-distribution (six degree of freedom).

Model 3: GARCH (0, 1) with average of mean squared errors (MSE) over the last 10 years and normal distribution.

Model 4: GARCH (0, 1) with average of mean squared errors (MSE) over the last 10 years and *t*-distribution (six degree of freedom).

Model 5: Power GARCH (1, 1) with normal distribution.

Model 6: Power GARCH (1, 1) with *t*-distribution (six degrees of freedom).

Model 7: Power GARCH (0, 1) with average of root mean squared errors (RMSE) over the last 10 years and normal distribution.

Model 8: Power GARCH (0, 1) with average of root mean squared errors (RMSE) over the last 10 years and *t*-distribution (six degrees of freedom).

<sup>24</sup> We also estimated Models 1, 2, 5 and 6 during 1969–2007. We find that the generated uncertainty according to these four models cannot beat the disagreement alone to match the survey measure of uncertainty when we include the unstable period.

squared error in the mean forecast, the uncertainty from GARCH-type models is less volatile and thus matches better the survey measure of uncertainty. This underscores the important point that *ex ante* uncertainty has to be generated conditionally based on the information known to survey respondents when making their forecasts, which is exactly what GARCH-type models do. We should, however, note that the error-based measures of uncertainty including GARCH have failed to signal the slowly creeping uncertainty in inflation and output growth forecasts since 2002, as indicated by the density forecasts. This is because the corresponding forecast errors have continued to be small despite the slow but steady increase in uncertainty due to unusual financial market developments and political instability in recent years. Uncertainty estimates based on density forecasts have an obvious advantage in this regard.

#### 4. CONCLUDING REMARKS

Owing to the ready availability of point forecasts, disagreement among forecasters has been widely used as a proxy for aggregate uncertainty in the economics, accounting and finance literature. Lacking a theoretical basis, empirical evidence has been mixed as to whether the disagreement is a reliable measure for uncertainty. Using a standard decomposition of forecast errors into common and idiosyncratic shocks in a panel data setting, our paper demonstrates that under certain regularity conditions the difference between uncertainty and disagreement is the perceived variance of future aggregate shocks that accumulate over forecast horizons. This result has important implications. It implies that the robustness of the relationship between uncertainty and disagreement depends on the variance of aggregate shocks over time and across horizons. Using the SPF density forecasts for inflation and output growth, we find direct evidence in support of our hypothesized time and horizon effects. As for the time effect, disagreement is found to be a reliable measure for uncertainty in a stable period. In periods with large volatility of aggregate shocks, however, disagreement becomes less useful as a proxy. As for the horizon effect, we find that the longer the forecast horizon, the larger is the difference between disagreement and uncertainty. Though disagreement alone tends to understate the level of uncertainty, our empirical results suggest that one can safely use disagreement as a proxy for uncertainty in a regression context, provided the forecast environment is relatively stable. By subtracting observed disagreement from uncertainty using density forecasts, we obtain a truly *ex ante* measure of aggregate shocks that befell the economy. These aggregate shocks are available to a policy maker before the actual values are realized, and show remarkable reduction in volatility after 1991.

Our results do not support the use of squared mean forecast errors to construct *ex ante* uncertainty, as often practiced in recent accounting and finance research. Since forecast error is an *ex post* measure reflecting unexpected shocks after the forecast is made, it should not affect uncertainty at the time of forecast. In order to construct an *ex ante* measure of forecast uncertainty, one should use the sum of the observed disagreement from the survey and the projected variance of aggregate shocks generated by a suitably specified GARCH-type model. We find that this approach performs much better than the use of squared forecast errors in matching the survey measure of uncertainty, and is less sensitive to occasional large forecast surprises. However, we should point out that still there remains some discrepancy between the survey measure and error-based *ex ante* measures of uncertainty that include the GARCH estimates. Further research on refining the GARCH specification may help.

## ACKNOWLEDGEMENTS

An earlier version of this paper was circulated as IFO Working Paper No. 60, and was presented at KOF Swiss Economic Institute and the Central Bank of Hungary. We thank Kai Carstensen, Antony Davies, Rob Engle, David Hendry, Gernot Nerb, Paul Söderlind, Jan-Egbert Sturm, Ken Wallis, Victor Zarnowitz, the guest editor Matteo Ciccarelli and an anonymous referee for many helpful comments and suggestions. However, we alone are responsible for any remaining errors and shortcomings.

## REFERENCES

- Abarbanell JS, Lanen WN, Verrecchia RE. 1995. Analysts' forecasts as proxies for investor beliefs in empirical research. *Journal of Accounting and Economics* **20**: 31–60.
- Banerjee A, Marcellino M. 2006. Are there any reliable leading indicators for US inflation and GDP growth? *International Journal of Forecasting* **22**: 137–151.
- Barron OE, Kim O, Lim SC, Stevens DE. 1998. Using analysts' forecasts to measure properties of analysts' information environment. *Accounting Review* **73**: 421–433.
- Barry CB, Jennings RH. 1992. Information and diversity of analyst opinion. *Journal of Financial and Quantitative Analysis* **27**: 169–183.
- Batchelor R, Dua P. 1993. Survey vs. ARCH measures of inflation uncertainty. *Oxford Bulletin of Economics and Statistics* **55**: 341–353.
- Batchelor R, Dua P. 1995. Forecast diversity and the benefits of combining forecasts. *Management Science* **41**: 68–75.
- Batchelor R, Dua P. 1996. Empirical measures of inflation uncertainty: a cautionary note. *Applied Economics* **28**: 333–341.
- Bernanke BS. 2007. Federal Reserve Communications. Speech delivered at the Cato Institute 25th Annual Monetary Conference, Washington, 14 November.
- Boero G, Smith J, Wallis KF. 2008. Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *Economic Journal* **118**: 1107–1127.
- Bollerslev T, Wooldridge JM. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews* **11**: 143–172.
- Bomberger WA. 1996. Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking* **28**: 381–392.
- Bomberger WA. 1999. Disagreement and uncertainty: a reply to Rich and Butler. *Journal of Money, Credit and Banking* **31**: 273–276.
- Campbell SD. 2007. Macroeconomic volatility, predictability, and uncertainty in the great moderation: evidence from the Survey of Professional Forecasters. *Journal of Business and Economic Statistics* **25**: 191–200.
- Davies A, Lahiri K. 1995. A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics* **68**: 205–227.
- Davies A, Lahiri K. 1999. Re-examining the rational expectations hypothesis using panel data on multiperiod forecasts. In *Analysis of Panels and Limited Dependent Variable Models*, Hsiao C, Pesaran MH, Lahiri K, Lee LF (eds). Cambridge University Press: Cambridge, UK; pp 76–90.
- Diebold FX, Tay AS, Wallis KF. 1999. Evaluating density forecasts of inflation: the survey of professional forecasters. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, Engle RF, White H (eds). Oxford University Press: Oxford; pp 76–90.
- Diether KB, Malloy CJ, Scherbian A. 2002. Differences of opinion and the cross section of stock returns. *Journal of Finance* **57**: 2113–2141.
- Engelberg J, Manski CF, Williams J. 2009. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics* **165**: 146–158.
- Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1008.



- Engle RF. 1983. Estimates of the variance of U.S. inflation based upon the ARCH model. *Journal of Money, Credit and Banking* **15**: 286–301.
- Giordani P, Söderlind P. 2003. Inflation forecast uncertainty. *European Economic Review* **47**: 1037–1059.
- Harvey D, Newbold P. 2003. The non-normality of some macroeconomic forecast errors. *International Journal of Forecasting* **19**: 635–653.
- Johnson TC. 2004. Forecast dispersion and the cross section of expected returns. *Journal of Finance* **59**: 1957–1978.
- Lahiri K, Liu F. 2005. ARCH models for multi-period forecast uncertainty: a reality check using a panel of density forecasts. In *Advances in Econometrics*. Vol. 20: *Econometric Analysis of Financial Time Series*, Fomby T, Terrell D (eds). Elsevier: Amsterdam; 321–363.
- Lahiri K, Liu F. 2006. Modeling multi-period inflation uncertainty using a panel of density forecasts. *Journal of Applied Econometrics* **21**: 1199–1220.
- Lahiri K, Sheng X. 2008. Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics* **144**: 325–340.
- Lahiri K, Sheng X. 2010. Learning and heterogeneity in GDP and inflation forecasts. *International Journal of Forecasting*, doi:10.1016/j.ijforecast.2009.12.009.
- Lahiri K, Teigland C, Zaporowski M. 1988. Interest rates and the subjective probability distribution of inflation forecasts. *Journal of Money, Credit and Banking* **20**: 233–248.
- Mankiw NG, Reis R, Wolfers J. 2003. Disagreement about inflation expectations. In *Macroeconomics Annual*, Gertler M, Rogoff K (eds). MIT Press: Cambridge, MA; 209–248.
- Markowitz HM. 1959. *Portfolio Selection: Efficient Diversification of Investments*. Wiley: New York.
- McConnell MM, Perez-Quiros G. 2000. Output fluctuations in the United States: what has changed since the early 1980's? *American Economic Review* **90**: 1464–1476.
- Mishkin FS. 2007. Inflation dynamics. Speech delivered at the Annual Macro Conference, Federal Reserve Bank of San Francisco, CA, 23 March.
- Mohanram PS, Sunder SV. 2006. How has regulation FD affected the operations of financial analysts? *Contemporary Accounting Research* **23**: 491–525.
- Öller LE, Barot B. 2000. The accuracy of European growth and inflation forecasts. *International Journal of Forecasting* **16**: 293–315.
- Patton A, Timmermann A. 2008. *Learning in real time: theory and empirical evidence from the term structure of survey forecasts*. Working paper, University of California, San Diego.
- Reifschneider D, Tulip P. 2007. *Gauging the uncertainty of the economic outlook from historical forecasting errors*. Finance and Economics Discussion Series 60, Federal Reserve Board, Washington, DC.
- Rich RW, Butler JS. 1998. Disagreement as a measure of uncertainty: a comment on Bomberger. *Journal of Money, Credit and Banking* **30**: 411–419.
- Stock JH, Watson MW. 2007. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking* **39**: 3–33.
- Wallis KF. 2004. An assessment of Bank of England and National Institute inflation forecast uncertainties. *National Institute Economic Review* no. (189): 64–71.
- Wallis KF. 2005. Combining density and interval forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics* **67**: 983–994.
- Zarnowitz V, Lambros LA. 1987. Consensus and uncertainty in economic prediction. *Journal of Political Economy* **95**: 591–621.
- Zhang XF. 2006. Information uncertainty and stock returns. *Journal of Finance* **61**: 105–137.