A comparative assessment of alternative ex ante measures of inflation uncertainty

Matthias Hartmann,* Helmut Herwartz§ and Maren Ulm‡

Abstract

Since the onset of the financial- and sovereign debt crisis in 2008, unconventional monetary and fiscal policy arrangements in industrialized economies have been raising concerns about the future evolution of inflation rates. The question of how to quantify inflation uncertainty, however, is an open issue. To assess the informative content of alternative ex ante quantifications of inflation uncertainty, we predict ex post squared inflation forecast errors in an out-of-sample forecasting contest. We find that the average across distinct models' ex ante uncertainty offers higher predictive power in comparison with other uncertainty measures based on the cross sectional variance of point forecasts, GARCH- or stochastic volatility models.

JEL classification: C52, C53, E31, E43
Keywords: Uncertainty measurement, ex ante and ex post inflation uncertainty, disagreement, out-of-sample forecasting, economic crisis.

*Corresponding author. Ruprecht-Karls-University Heidelberg, Alfred-Weber-Institute for Economics, Bergheimer Strasse 58, 69115 Heidelberg, Germany. Tel.: +49 621 54 2908; Fax: +49 621 54 3649; email: matthias.hartmann@awi.uni-heidelberg.de.

§Georg-August-University Goettingen, Department of Economics. email: hherwartz@uni-goettingen.de.

‡Georg-August-University Goettingen, Department of Economics. email: maren.ulm@wiwi.uni-goettingen.de.

Financial support by the Deutsche Forschungsgemeinschaft (Grant No. HE 2188/3-3) is gratefully acknowledged.

1 Introduction

Since the outbreak of the financial- and sovereign debt crisis, the effects of macroeconomic uncertainty have been documented in a growing theoretical and empirical literature. For example, Bloom (2009) or Jurado et al. (2015) investigate the influence of uncertainty on real economic activity. Bloom (2009) explains this linkage through investment decisions of firms, which can be affected by the uncertainty regarding the future payoffs of physical investment projects. Early contributions with an explicit focus on uncertainty about future inflation are Okun (1971) and Friedman (1977). Friedman (1977) highlights the detrimental effects of inflation uncertainty (IU in the following) on aggregate investment and output. One reason for the sustained interest in IU might be the ongoing dispute about the sources of the so-called Great Moderation. The Great Moderation describes a secular containment of inflation fluctuations that has been observed through recent decades across industrialised economies (McConnel and Perez-Quiros 2000, Benati 2008, Herrera and Pesavento 2009, Lahiri and Sheng, 2010). However, empirical studies on the causes and effects of IU face the problem that IU is unobservable.

The aim of this study is to evaluate a broad range of currently employed measures of IU. Conceptually, most IU statistics are either derived from dynamic specifications like, e.g., (G)ARCH and stochastic volatility (SV) models, or from the information provided by forecast surveys (Zarnowitz and Lambros 1987, Giordani and Söderlind 2003, Lahiri and Sheng 2010). Representatives of the former category draw upon the historical time series information. In survey-based approaches, in contrast, IU is often approximated by the cross sectional dispersion of point forecasts or by the average over survey participants' individual uncertainty as it is derived from density forecasts (Giordani and Söderlind 2003, Rich and Tracy 2003). The high predictive content of survey based point forecasts for inflation is documented, e.g., by Ang et al. (2007). Hence, this approach seems also promising as a means to quantify IU. Clements (2014) compares measures of ex ante forecast uncertainties derived from survey based density forecasts of inflation and output growth with ex post, i.e. realised, uncertainties that are derived from forecast errors. The magnitude of ex-ante IU as compared to ex-post IU is interpreted as a metric of over- or underconfidence. Clements (2014) finds that ex ante and ex post uncertainty are only weakly related, in particular at short forecast horizons. Since survey- and time series methods aggregate information in distinct ways (Batchelor and Dua 1996; Mankiw and Reis 2004), they will often provide diverging estimates of IU. Moreover, as argued by Lahiri and Sheng (2010), choosing between these two approaches might become most difficult during turbulent periods.

In this study, IU statistics from both the survey- and the time series category are assessed according to their performance as predictor variables. Similar to Clements (2014), ex ante IU is related to ex post IU. The ex post, i.e. realised, IU obtains as squared inflation forecast errors. We then forecast ex post IU by alternative ex ante IU measures and rank them in terms of their predictive content. To improve comparability between time series and disparity type IU statistics, inflation expectations from surveyed experts are represented by the predictions of various econometric forecasting models. The first advantage of this approach is that differences in the predictive content cannot arise due to distinct information sets. Both the time series based approaches and the dispersion statistics rely on the same historical time series information. Second, the consideration of a larger cross section of economies is facilitated, since for sufficiently long time periods, survey data on IU are only available for the Euro area and the US. Similarly, Brock and Hommes (1997) or Branch (2004) model the heterogeneity of expectations in terms of a finite number of prediction models. In contrast to other studies which evaluate alternative IU statistics, our investigation is based on a large scale international data set, covering 18 industrialised economies and the sample period between 1997 and 2014. This allows to compare the features and relative performance of IU measures during the time after the onset of the financial- and sovereign debt crisis in 2007 as well as during the less turbulent period before.

We find that IU predictions which are based on the average uncertainty of alternative models are most accurate to predict ex post squared inflation forecast errors. In a descriptive analysis of the proposed IU statistics, correlation coefficients reveal that IU estimates belonging to either the time series- or the dispersion category process information in a similar way. However, each quantification of IU also shows idiosyncratic characteristics, which might explain the superior predictive performance of the average uncertainty statistic. Moreover, to enable the comparison with the model based IU statistics, we compute the disagreement among forecasters from the *Consensus Economics* survey with regard to their inflation expectations. The disagreement derived from *Consensus Economics* data is positively correlated with the model based disagreement and the average uncertainty implied by the individual models. For all IU quantifications, a strong increase of uncertainty is observed at the breakout of the global financial crisis. Interestingly, we do not find evidence for a substantial decline of IU after 2009 in high inflation economies, as it is the case for countries characterized by relatively low inflation rates.

In Section 2 we introduce six competing IU metrics. In Section 3 we describe the relationship between ex ante and ex post IU that is employed to assess the predictive content of alternative IU statistics. Moreover, the data set and the IU measures are described. An introduction of the forecasting design and the discussion of the forecast comparisons follows in Section 4. Section 5 summarises and concludes. The alternative forecasting models used to substitute survey forecasts are outlined in the Appendix.

2 Measuring IU

There is no unique way to define or quantify IU. Thus, we seek to determine the ex ante IU statistic that has the highest predictive content among a set of alternatives that have been proposed in the literature. We evaluate IU measures which mimic commonly used dynamic and disparity approaches. Following, e.g., Hamilton (1985), Brock and Hommes (1997) or Branch (2004), we compute measures of disparity by substituting survey expectations with forecasts that are derived from a variety of econometric (time series) models. This procedure allows to analyse a larger cross section of economies, since typical forecast surveys such as the *Survey of Professional Forecasters* are only available for the US or the Euro area. Moreover, this approach guarantees an equal timing of information sets underlying both time series based and disparity type IU measures. In the following, six distinct IU statistics are reviewed. We firstly consider time series based methods, which include GARCH and stochastic volatility (SV) models, and subsequently illustrate approaches which are based on the dispersion of individual forecasts. All measures are ex ante quantifications of IU. Several of the IU statistics discussed below are based (at least partly) on the linear autoregressive (AR) model, a frequently used specification for inflation forecasting. The comparably strong predictive performance of AR and random walk specifications for inflation processes is documented in several empirical studies, including Canova (2007) or Stock and Watson (2007, 2008). The AR scheme is formulated as

$$\pi_{t+\ell} = \mu + \alpha_{11}(L)\pi_t + \varepsilon_{t+\ell}, \quad t = \tau - B + 1, ..., \tau. \tag{1}$$

In (1), $\varepsilon_{t+\ell} \sim (0, \sigma_\varepsilon^2)$, L denotes the lag operator, i.e. $L^n\pi_t = \pi_{t-n}$, and $\alpha_{11}(L) = \alpha_{11,0} + \alpha_{11,1}L + \cdots + \alpha_{11,P}L^P$. The lag order P is selected by means of the AIC, with maximum order set to $P^{\max} = 12$. We consider the forecast horizons $\ell \in \{1, 3, 6, 12\}$ and alternative lengths of a (rolling) estimation sample $B \in \{72, 108\}$. Out-of-sample forecasts implied by (1) are denoted $\hat{\pi}_{\tau+\ell|\tau}$, where $\tau = T_0 - \ell - P, ..., T - \ell$ is the rolling forecast origin. The time instances T_0 and T delimit the evaluation sample which is employed in the comparative evaluation of IU measures. Note that the estimation of the dynamic model in (1) requires $P^{\max}$ 'presample' observations.

1. Measuring IU by means of dynamic specifications

1.1 Predictive standard deviation

At forecast origin τ, the estimated predictive error standard deviation is

$$\hat{\sigma}_{\tau+\ell|\tau} = \tilde{\sigma}_{\epsilon}\sqrt{(1 + \boldsymbol{z}'_{\tau}(Z'_{\tau}Z_{\tau})^{-1}\boldsymbol{z}_{\tau})}, \tag{2}$$

where $\tilde{\sigma}_{\varepsilon}$ is the OLS estimator of σ_{ε} in (1), Z_{τ} is the autoregressive design matrix and $\boldsymbol{z}_{\tau}$ are the most recent observations on which out-of-sample forecasts are based. The statistic in (2) is composed of time-local expressions of the variance of inflation surprises and estimation uncertainty.

1.2 GARCH(1,1)

Since the studies of Engle (1982, 1983) and Bollerslev (1986), (G)ARCH models have been widely used to measure IU. In this framework, the conditional variance of the inflation disturbances filtered by the autoregressive scheme in (1) reads as

$$\hat{h}_t^2 = \omega + \gamma\varepsilon_{t-1}^2 + \beta\hat{h}_{t-1}^2, \tag{3}$$

with $\omega > 0,\ \gamma > 0,\ \beta > 0$, and $\gamma + \beta < 1$. The GARCH(1,1) specification has been found to be particularly suitable for out-of-sample forecasting (cf. Akgiray 1989, Knight and Satchell 2007). The coefficients in (3) are estimated by means of (quasi-) maximum likelihood estimation within rolling windows of fixed size $B \in \{72, 108\}$. Out-of-sample forecasts of the conditional volatility of the filtered residuals can be obtained as

$$\hat{h}_{G,\tau+\ell|\tau} = \sqrt{\overline{h^2} + (\gamma+\beta)^{\ell-1}(\hat{h}_{\tau+1}^2 - \overline{h^2})}, \tag{4}$$

where $\mathbf{E}[h_t^2] := \overline{h^2} = \omega/(1-\gamma-\beta)$.

1.3 Unobserved components stochastic volatility (UCSV)

According to the UCSV model (Stock and Watson, 2007), IU is modeled as the sum of uncertainties induced by a transitory shock η_t on the observed inflation rate π_t and a shock ϵ_t affecting the trend inflation ϖ_t. The UCSV model is specified as

$$\pi_{t+1} = \varpi_{t+1} + \eta_{t+1}, \quad \eta_{t+1} = \sigma_{\eta,t+1}\zeta_{\eta,t+1},$$
$$\varpi_{t+1} = \varpi_t + \epsilon_{t+1}, \quad \epsilon_{t+1} = \sigma_{\epsilon,t+1}\zeta_{\epsilon,t+1}.$$

By assumption, the conditional log variances follow independent random walks, i.e.

$$\ln \sigma^2_{\eta,t+1} = \ln \sigma^2_{\eta,t} + \nu_{\eta,t+1} \text{ and } \ln \sigma^2_{\epsilon,t+1} = \ln \sigma^2_{\epsilon,t} + \nu_{\epsilon,t+1}.$$

The variances $\sigma^2_{\eta,t+1}$ and $\sigma^2_{\epsilon,t+1}$ represent the transitory and the permanent component of inflation fluctuations, respectively. The innovations $\zeta_{t+1} = (\zeta_{\eta,t+1}, \zeta_{\epsilon,t+1})$ are assumed to satisfy $N(0, I_2)$. Moreover, we presume that $\nu_{t+1} = (\nu_{\eta,t+1}, \nu_{\epsilon,t+1}) \sim N(0, \theta I_2)$, where the parameter θ governs the smoothness of the variance processes. The error terms $\zeta_{\tau+1}$ and $\nu_{\tau+1}$ are assumed to be mutually independent. The UCSV model is estimated by means of Gibbs sampling. The variance parameter is set to $\theta = (0.2/3)^2$ following the specification in Dovern et al. (2012). We assess the overall uncertainty about inflation as

$$\hat{h}_{SV,\tau+\ell|\tau} = \sqrt{\sigma^2_{\eta,\tau+\ell|\tau} + \sigma^2_{\epsilon,\tau+\ell|\tau}}. \quad (5)$$

Forecasts of $\hat{h}_{SV,\tau+\ell|\tau}$ for $\ell > 1$ are calculated similar to the ex ante forecast uncertainty for the UCSV Model in Clements and Galvão (2014).[1]

2. Measuring IU by means of disparity

A common way of measuring uncertainty is to exploit the variation across individual forecasts. We consider a range of forecasts from $J = 14$ alternative model specifications, namely the AR scheme in (1) and an extended version of (1) obtained by incorporating the deviation of inflation from its long run trend. Moreover, an adaptive expectations specification proposed by Branch (2004), and a set of distinct vector autoregressive (VAR) schemes is incorporated. These forecasting models are listed in the Appendix.

2.1 Disagreement among expectations

In the absence of density forecasts or statements about individual uncertainty of surveyed experts, the disagreement among forecasts or expectations is often used as an indicator of uncertainty (Rich et al. 2012). Based on rival predictions of inflation, denoted

[1] However, in our case there is no need to include the weighting factor that is adopted by Clements and Galvão (2014) for annual data.

$\hat{\pi}_{j,\tau+\ell|\tau}$, $j = 1, ..., J$, with $J = 14$, the disagreement obtains as

$$\hat{s}_{\tau+\ell|\tau} = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}(\hat{\pi}_{j,\tau+\ell|\tau} - \overline{\pi}_{\tau+\ell|\tau})^2}, \tag{6}$$

where $\overline{\pi}_{\tau+\ell|\tau} = (1/J)\sum_{j=1}^{J}\hat{\pi}_{j,\tau+\ell|\tau}$. Mankiw and Reis (2004) argue that disagreement about inflation expectations plays a crucial role for macroeconomic dynamics. Capistrán and Ramos-Francia (2010) or Dovern et al. (2012) investigate the determinants of disagreement among forecasters' inflation expectations. Diether et al. (2002) or Yu (2011) explain the variation of a cross section of stock returns by means of the disagreement among financial market analysts. Mokinski et al. (2015) provide a comprehensive evaluation of alternative ways to derive disagreement statistics based on qualitative and quantitative expectations.

2.2 Average uncertainty

The popularity of the disagreement statistic in (6) might be partly due to the fact that it only requires the availability of point forecasts. However, disagreement as such is difficult to interpret. First, it is not directly linked to idiosyncratic uncertainty. Second, disagreement may be only one component of aggregate uncertainty, e.g., if the latter is derived from a set of density forecasts (Lahiri et al. 1988; Lahiri and Liu 2005; Wallis 2005; Boero et al. 2008). Empirically, D'Amico and Orphanides (2008) and Rich et al. (2012) find only a weak relation between the disagreement obtained from the *Survey of Professional Forecasters* and aggregate IU. As an alternative to a disagreement statistic akin to (6), Zarnowitz and Lambros (1987) propose to average individual predictive standard deviations to quantify IU. Such a metric is given by

$$\bar{\sigma}_{\tau+\ell|\tau} = \frac{1}{J}\sum_{j=1}^{J}\hat{\sigma}_{j,\tau+\ell|\tau}, \tag{7}$$

with $\hat{\sigma}_{j,\tau+\ell|\tau}$ denoting model specific predictive standard deviations obtained, e.g., according to (2) for the AR scheme in (1). The statistic $\bar{\sigma}_{\tau+\ell|\tau}$ is regarded as a dispersion measure like $\hat{s}_{\tau+\ell|\tau}$, since both entail characteristics which only arise as a matter of pooling. However, as a forecast combination, $\bar{\sigma}_{\tau+\ell|\tau}$ is less likely to obtain an 'eccentric' assessment of IU than each of its individual components (Zarnowitz and Lambros 1987).

2.3 Joint uncertainty

The IU statistics $\bar{\sigma}_{\tau+\ell|\tau}$ and $\hat{s}_{\tau+\ell|\tau}$ may be interpreted as two components of the aggregated IU of a cross section of forecasters (Lahiri et al. 1988). Moreover, Lahiri et al. (2014) show that a quantification of the aggregate uncertainty of a combined forecast should incorporate both the average individual uncertainty and the disagreement of point predictors. Therefore, it appears tempting to join the measures described in (6) and (7). Forecast combination strategies for conditional second order moments have been evaluated by Becker and Clements (2008) or Patton and Sheppard (2009), who investigate volatility forecasts for the S&P500 index and IBM stock returns, respectively. In both cases, averages of single model based volatility forecasts are found to have higher predictive accuracy than competing prediction schemes. The combined ex ante uncertainty statistic reads as

$$\varsigma_{\tau+\ell|\tau} = 0.5(\hat{s}_{\tau+\ell|\tau} + \bar{\sigma}_{\tau+\ell|\tau}). \tag{8}$$

3 Empirical evaluation of inflation uncertainty

The evaluation of alternative IU quantifications proceeds by means of a forecast comparison which ranks IU statistics according to their predictive content. Similar to Clements (2014), we compare ex ante IU measures with ex post IU that reads as

$$e^2_{\tau+\ell} \quad = \quad (\pi_{\tau+\ell} - \bar{\hat{\pi}}_{\tau+\ell|\tau})^2, \tag{9}$$

where $\bar{\hat{\pi}}_{\tau+\ell|\tau} = (1/J)\sum_{j=1}^{J} \hat{\pi}_{j,\tau+\ell|\tau}$ is a forecast combination based on inflation forecasts from (1) and the models (15) to (19) that are outlined in the Appendix. To determine the IU measure that delivers the highest predictive content, we employ the alternative IU statistics as predictor variables. The relation to assess IU metrics reads as

$$\ln(e^2_{\tau+\ell}) \quad = \quad \delta_0 + \delta_1(L)\ln(e^2_\tau) + \delta_2(L) IU^\bullet_{\tau+\ell|\tau} + \vartheta^\bullet_{\tau+\ell}, \quad \vartheta^\bullet_{\tau+\ell} \sim (0, \sigma^2_\vartheta), \tag{10}$$

where the log transformation in (10) ensures positiveness of predictions for $e^2_{\tau+\ell}$. Moreover, $\tau = T_0 - \ell, ..., T - \ell$ and $\delta_k(L) = \delta_{k,0} + \delta_{k,1}L + \cdots + \delta_{k,P}L^P,\ k = 1, 2$, and $P^{\max} = 12$. The '$\bullet$' notation indicates that $IU^\bullet_{\tau+\ell|\tau} \in \{\hat{\sigma}_{\tau+\ell|\tau}, \hat{h}_{G,\tau+\ell|\tau}, \hat{h}_{SV,\tau+\ell|\tau}, \hat{s}_{\tau+\ell|\tau}, \bar{\sigma}_{\tau+\ell|\tau}, \varsigma_{\tau+\ell|\tau}\}$ represents one out of the alternative IU measures listed in (2) to (8). To evaluate the improvement in prediction accuracy that is provided by single IU estimates, forecasts from (10) are compared with the ones where IU is omitted from (10), i.e. for

$\delta_{2,0} = \delta_{2,1} = ... = \delta_{2,P} = 0$.

3.1 Data

The data set comprises monthly observations for 18 developed economies, namely Austria (AU), Belgium (BE), Canada (CA), Denmark (DK), Finland (FI), France (FR), Germany (GER), Ireland (IR), Italy (IT), Japan (JP), the Netherlands (NL), Norway (NOR), Portugal (PT), Spain (SP), Sweden (SWE), Switzerland (SWI), the UK and the US for the time period January 1997 to December 2014. Based on consumer prices, denoted CPI_t, we obtain inflation series as monthly year-on-year price changes, i.e. $\pi_t = \ln(CPI_t/CPI_{t-12})$. Similarly, real activity as the year-on-year change in the industrial production index is determined as $\tilde{y}_t = \ln(IP_t/IP_{t-12})$. Further data series that are employed in prediction models are described in the Appendix. These are the respective economies' central bank target rate, R_t, oil prices in local currency, oil_t, a monetary aggregate, m_t, and the unemployment rate $unemp_t$. All series are seasonally adjusted and drawn from Datastream.

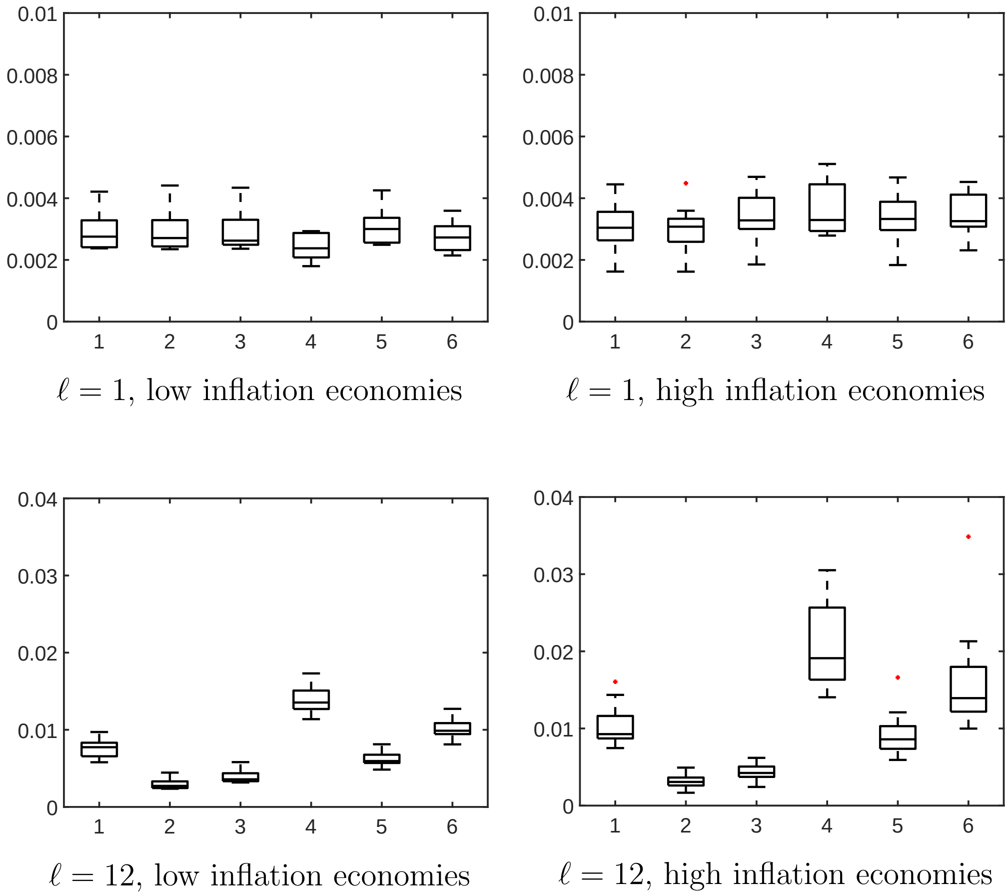
3.2 Descriptive analysis of IU measures

Before assessing the predictive content of the competing IU estimates, descriptive features like their relative magnitudes, mutual correlations and graphical displays are examined. The boxplots in Figure 1 show the size and variation of the IU approximations for the cross section of 18 economies, separated into two groups of high[2] and low[3] inflation economies. This distinction is drawn by separating economies with average inflation rates above and below the cross sectional median. As it turns out, the relative magnitudes of the IU statistics considered in this study are similar to those reported by Batchelor and Dua (1996) or Bomberger (1996). The displays for $\ell = 1$ depict averages of $IU_{\tau+1|\tau}$ over the period between January 1997 and December 2014 for the full cross section. For this horizon, we observe only minor variations among IU measures. In case of the high inflation economies, the variation is slightly larger for dispersion statistics in comparison with time series based IU metrics. Both the size and the variation of average IU estimates across economies increase for $\ell = 12$ compared with $\ell = 1$. The increase is most notable for the disagreement statistic $\hat{s}_{\tau+\ell|\tau}$. In contrast to the short horizon statistics ($\ell = 1$), there is a remarkable difference in the size and variation of the

[2]High inflation economies: FI, IR, IT, NOR, PT, SP, SWE, UK, US.

[3]Low inflation economies: AU, BE, CA, DK, FR, GER, JP, NL, SWI.

dispersion statistics between low and high inflation economies. In contrast, time series based IU statistics differ markedly less across low and high inflation economies.



$\ell = 1$, low inflation economies

$\ell = 1$, high inflation economies

$\ell = 12$, low inflation economies

$\ell = 12$, high inflation economies

Figure 1: Boxplots for average IU over the time period 1997M1 to 2014M12 in low and high inflation economies. The numbers from 1.) to 6.) refer to the following IU measures: 1.) $\hat{\sigma}_{\tau+\ell|\tau}$, 2.) $\hat{h}_{G,\tau+\ell|\tau}$, 3.) $\hat{h}_{SV,\tau+\ell|\tau}$, 4.) $\hat{s}_{\tau+\ell|\tau}$, 5.) $\bar{\sigma}_{\tau+\ell|\tau}$, 6.) $\varsigma_{\tau+\ell|\tau}$. The scales of IU graphs for $\ell = 1$ and $\ell = 12$ differ to facilitate the visual inspection of variation in the IU statistics.

We further assess the similarity between the model based disagreement statistics and a commonly used survey-based estimate of IU. To determine the survey-based measure, we make use of the data set provided by *Consensus Economics*. This monthly survey of expert forecasters elicits inflation expectations for the current and the next calendar year for a subset of the countries and the time span we consider.[4]

[4]The reduced sample consists of CA, FR, GER, IT, JP, NL, NOR, SP, SWI, the UK and the US,

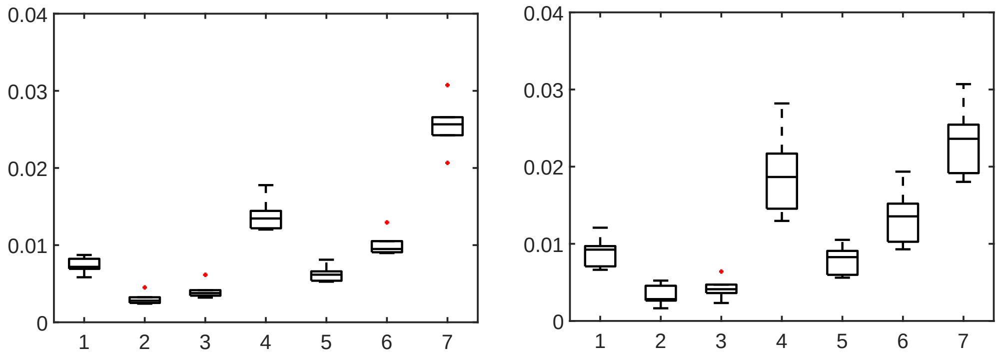
To facilitate the comparison with the fixed horizon IU statistics in (2) to (8), we transform the expectations from the *Consensus Economics* data set. The rolling horizon structure of the survey expectations is converted by means of a weighted average of predictions for the current and the next year. This approximation should resemble the fixed horizon structure more closely and is, e.g., suggested by Dovern et al. (2012). The disagreement of $j = 1, ..., \mathcal{J}$ *Consensus Economics* forecasts is obtained in analogy to $\hat{s}_{\tau+\ell|\tau}$ in (6) as

$$u_{\tau+12|\tau} = \sqrt{\frac{1}{\mathcal{J}-1}\sum_{j=1}^{\mathcal{J}}(\hat{\pi}_{j,\tau+12|\tau} - \overline{\pi}_{\tau+12|\tau})^2}, \tag{11}$$

with $\overline{\pi}_{\tau+12|\tau} = (1/\mathcal{J})\sum_{j=1}^{\mathcal{J}}\hat{\pi}_{j,\tau+12|\tau}$. The number of available inflation forecasts varies across time instances and economies. On average, $u_{\tau+12|\tau}$ is estimated on the basis of approximately 20 forecasts.

Figure 2 shows the magnitude and variation of the alternative IU quantifications including the disagreement statistic $u_{\tau+12|\tau}$ for $\ell = 12$. To ensure comparability, the calculations are based on the reduced sample for all IU statistics. For the low inflation economies, the magnitude of IU estimate $u_{\tau+12|\tau}$ is higher than for the model based IU metrics. The amount of variation of $u_{\tau+12|\tau}$ is similar to the disagreement measure $\hat{s}_{\tau+\ell|\tau}$, especially for the low inflation economies. For the remaining IU metrics, both the size and the cross sectional variation are larger for the disparity measures than for the time series quantifications. The considerable differences across IU approximations exemplify that they supply idiosyncratic information.

for the period from 1999M6 to 2014M12.



$\ell = 12$, low inflation economies $\qquad$ $\ell = 12$, high inflation economies

Figure 2: Boxplots for average IU over the time period 1999M6 to 2014M12 in low and high inflation economies for $\ell = 12$. The numbers from 1.) to 7.) refer to the following IU measures: 1.) $\hat{\sigma}_{\tau+\ell|\tau}$, 2.) $\hat{h}_{G,\tau+\ell|\tau}$, 3.) $\hat{h}_{SV,\tau+\ell|\tau}$, 4.) $\hat{s}_{\tau+\ell|\tau}$, 5.) $\bar{\sigma}_{\tau+\ell|\tau}$, 6.) $\varsigma_{\tau+\ell|\tau}$, 7.) $u_{\tau+12|\tau}$

To provide an indication of the extent of common information in the IU metrics, we next examine their mutual correlations. As noted by Lahiri and Sheng (2010), the degree to which characteristics of IU measures differ might depend on the level of inflation. Therefore, the average correlation coefficients for the two horizons $\ell = 1, 12$ are reported separately for low and high inflation economies in Tables 1 and 2, respectively.

Table 1: Mutual correlations of IU measures at the short horizon

	$\hat{\sigma}_{\tau+1\|\tau}$	$\hat{h}_{G,\tau+1\|\tau}$	$\hat{h}_{SV,\tau+1\|\tau}$	$\hat{s}_{\tau+1\|\tau}$	$\bar{\sigma}_{\tau+1\|\tau}$
High inflation economies					
$\hat{h}_{G,\tau+1\|\tau}$	0.59	·	·	·	·
$\hat{h}_{SV,\tau+1\|\tau}$	0.41	0.72	·	·	·
$\hat{s}_{\tau+1\|\tau}$	0.57	0.35	0.38	·	·
$\bar{\sigma}_{\tau+1\|\tau}$	0.85	0.45	0.34	0.72	·
$\varsigma_{\tau+1\|\tau}$	0.68	0.41	0.40	0.98	0.84
Low inflation economies					
$\hat{h}_{G,\tau+1\|\tau}$	0.56	·	·	·	·
$h_{SV,\tau+1\|\tau}$	0.39	0.71	·	·	·
$\hat{s}_{\tau+1\|\tau}$	0.36	0.29	0.36	·	·
$\bar{\sigma}_{\tau+1\|\tau}$	0.77	0.53	0.47	0.56	·
$\varsigma_{\tau+1\|\tau}$	0.49	0.38	0.41	0.98	0.72

Note: The upper panel contains correlation coefficients between IU measures for nine high inflation economies, while the lower part refers to nine low inflation economies.

The overall impression for both high and low inflation economies is that mutual correlations are stronger within than between the two categories of IU quantifications. Particularly high correlations are found among the dispersion measures $\hat{s}_{\tau+\ell|\tau}$ and $\varsigma_{\tau+\ell|\tau}$. The disagreement statistic $u_{\tau+12|\tau}$ exhibits positive correlations especially with the disparity statistic $\bar{\sigma}_{\tau+\ell|\tau}$ and the dynamic measures $\hat{h}_{SV,\tau+\ell|\tau}$ and $\hat{h}_{G,\tau+1|\tau}$ (see Table 2).

Table 2: Mutual correlations of IU measures at the long horizon

	$\hat{\sigma}_{\tau+12\|\tau}$	$\hat{h}_{G,\tau+12\|\tau}$	$\hat{h}_{SV,\tau+12\|\tau}$	$\hat{s}_{\tau+12\|\tau}$	$\bar{\sigma}_{\tau+12\|\tau}$	$\varsigma_{\tau+12\|\tau}$
High inflation economies						
$\hat{h}_{G,\tau+12\|\tau}$	0.39	·	·	·	·	·
$\hat{h}_{SV,\tau+12\|\tau}$	0.26	0.56	·	·	·	·
$\hat{s}_{\tau+12\|\tau}$	0.47	0.31	0.43	·	·	·
$\bar{\sigma}_{\tau+12\|\tau}$	0.78	0.46	0.55	0.69	·	·
$\varsigma_{\tau+12\|\tau}$	0.56	0.36	0.46	0.99	0.80	·
$u_{\tau+12\|\tau}$	0.10	0.29	0.42	0.11	0.22	0.14
Low inflation economies						
$\hat{h}_{G,\tau+12\|\tau}$	−0.12	·	·	·	·	·
$\hat{h}_{SV,\tau+12\|\tau}$	−0.22	0.47	·	·	·	·
$\hat{s}_{\tau+12\|\tau}$	0.25	0.19	0.39	·	·	·
$\bar{\sigma}_{\tau+12\|\tau}$	0.44	0.16	0.40	0.64	·	·
$\varsigma_{\tau+12\|\tau}$	0.31	0.19	0.38	0.98	0.76	·
$u_{\tau+12\|\tau}$	−0.15	0.21	0.33	0.09	0.18	0.11

Note: The upper panel documents correlation coefficients between IU measures for six high inflation economies, while the lower part refers to the low inflation economies. Computations are based on the sample ranging from 1999M6 to 2014M12.

Figures 3 and 4 display the median of the IU realizations to illustrate their dynamic behaviour. At the short forecast horizon ($\ell = 1$, Figure 3), IU metrics differ only slightly between low and high inflation economies. These differences mainly emerge at the end of the 1990s and in the immediate aftermath of the global financial crisis around 2009. For all estimates the relatively stable evolution of IU before 2008 resembles the timing of the Great Moderation (Benati 2008). During the breakout of the global financial crisis all IU metrics except $\hat{\sigma}_{\tau+\ell|\tau}$ increase remarkably. This peak, however, is largely reversed before 2010. For $\hat{\sigma}_{\tau+\ell|\tau}$, in contrast, a minor rise is notable in 2008 which essentially persists until the end of the sample.

At the long horizon ($\ell = 12$, Figure 4), we diagnose substantial deviations of IU statistics between high and low inflation economies. The most sizeable deviations between high and low inflation economies are observed for the disparity measures $\hat{s}_{\tau+\ell|\tau}$, $\bar{\sigma}_{\tau+\ell|\tau}$, $\varsigma_{\tau+\ell|\tau}$ and the dynamic measure $\hat{\sigma}_{\tau+\ell|\tau}$. Following the substantial rise of IU after 2009, low inflation economies experience a reduction towards the pre-crisis level. For high inflation economies, IU decreases around 2012 but remains on an increased level

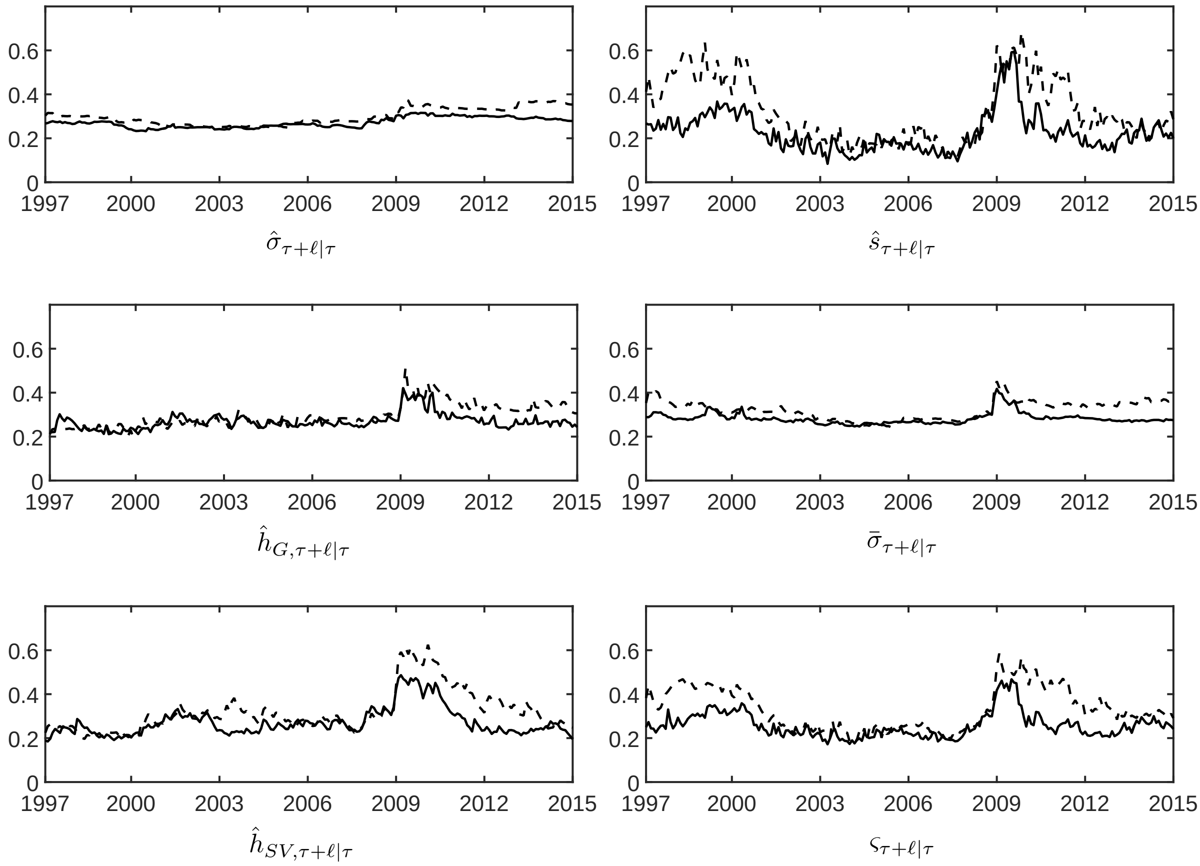


Figure 3: The graphs show the median of distinct IU approximations at the anticipation horizon $\ell = 1$ for low (solid line) and high inflation (dashed) economies. The scale of the IU measures has been multiplied by 100 in this figure to facilitate visual inspection.

until the end of the sample period.

Similarly, the *Consensus Economics* based disagreement $u_{\tau+12|\tau}$ displays a relatively stable evolution before 2008 and a strong peak for all economies at the breakout of the global financial crisis. For both high and low inflation economies and $\ell = 12$, upward shifts in IU are more pronounced for the disparity measures than for the time series approximations of IU.

The preliminary data analysis in this Section confirms that model based uncertainty statistics like $\hat{s}_{\tau+\ell|\tau}$ and the widely used survey based disagreement evolve in a broadly similar way. Moreover, the empirical correlations highlight the distinct features of time series vs. dispersion metrics of IU on the one hand, and the similarities within these groups on the other hand. This complementarity suggests that the quantification of IU might gain by combining time series and cross sectional information. Therefore, joint

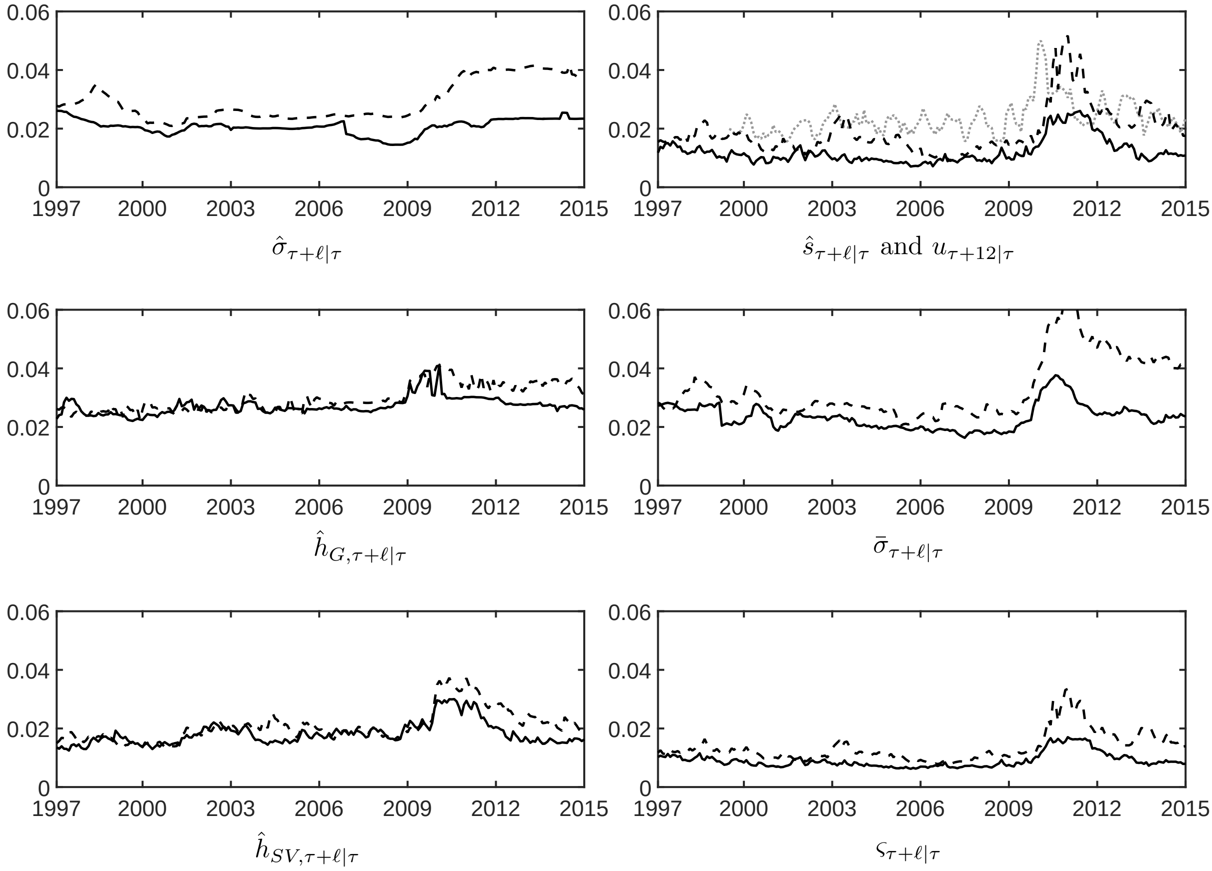


Figure 4: The graphs show the median of distinct IU approximations at the anticipation horizon $\ell = 12$. For further notes see Figure 2. The upper right graph for $\hat{s}_{\tau+\ell|\tau}$ additionally includes the trajectory for $u_{\tau+12|\tau}$ (grey dotted line). The IU measures have been individually rescaled to facilitate the visual inspection of their trajectories.

statistics such as $\bar{\sigma}_{\tau+\ell|\tau}$ or $\varsigma_{\tau+\ell|\tau}$ might offer advantages in terms of predictive content. To compare the relative merits of dynamic and dispersion IU metrics, we next devise an out-of-sample forecasting assessment.

4 The relation between ex ante and ex post IU

The comparative forecast evaluation of IU measures proceeds by means of a pseudo out-of-sample cross validation. Each realised squared forecast error, $\ln(e^2_{\tau+\ell}),\ \tau = T_0 - \ell, ..., T - \ell$, is predicted ℓ-steps ahead by means of a respective leave-one-out estimate. The computation of ℓ-steps ahead predictions is straightforward due to the linear relation between $\ln(e^2_{\tau+\ell})$ and the explanatory variables which are conditional on information that is available in period τ (Chevillon 2005). Subset modelling is implemented in terms of the Akaike information criterion (AIC) for each lag polynomial of the predictor variables in (10) and at each cross-validation step.[5]

4.1 The ranking of forecasts

From the predictions $\widehat{\ln(e^2)}^{\bullet}_{\tau+\ell|\tau}$ that are based on distinct IU quantifications included in (10) and the benchmark excluding IU, (i.e. $\delta_{2,0} = \delta_{2,1} = ... = \delta_{2,P} = 0$ in (10)), we derive rankings of forecast performance. The ranking schemes are introduced below. Loss statistics are aggregated across economies, hence, we introduce the index $i = 1, ..., N$, where $N = 18$.

4.1.1 Performance criteria

To assess the robustness of the out-of-sample findings, we report results for three distinct loss functions. Note that the forecasting model for ex post squared errors in (10) with distinct IU statistics and the respective specification without IU term are in most instances non-nested specifications due to the lag-selection step that is carried out before forecasts are computed. Thus, corresponding RMSE statistics can be compared by means of the Diebold-Mariano (DM) test statistic for $\ell = 1$ (Diebold and Mariano 1995) and the adjusted DM test (Harvey et al. 1997) for $\ell > 1$.

1. Root mean squared error (RMSE)

The RMSE of the prediction error from (10) ($\vartheta^{\bullet}_{i,\tau+\ell|\tau}$) reads as

$$\text{RMSE}^{\bullet} = \sqrt{(1/((T - T_0 + 1) \times 18)) \sum_{\tau=T_0-\ell}^{T-\ell} \sum_{i=1}^{18} \left(\vartheta^{\bullet}_{i,\tau+\ell|\tau}\right)^2}. \tag{12}$$

2. Directional accuracy (DA)

As a second criterion, we employ the DA loss function, given by

$$\text{DA}^{\bullet} = (1/((T - T_0 + 1) \times 18)) \sum_{\tau=T_0-\ell}^{T-\ell} \sum_{i=1}^{18} \text{I}((\hat{\pi}^{\bullet}_{i,\tau+\ell|\tau} - \pi_{i,\tau})(\pi_{i,\tau+\ell} - \pi_{i,\tau}) > 0), \tag{13}$$

[5] Results are qualitatively similar if the BIC is considered.

where $\mathrm{I}(\cdot)$ denotes an indicator function. Though $\mathrm{DA}^\bullet$ does not make use of the information regarding the size of $\vartheta^\bullet_{i,\tau+\ell|\tau}$, an advantage of $\mathrm{DA}^\bullet$ is its robustness with respect to the occurrence of extreme prediction errors. The close association with economic profits (Leitch and Tanner 1991, Swanson and White 1995) is a further advantage of DA. Blaskowitz and Herwartz (2014) provide a discussion of the economic value of directionally accurate predictions.

3. Top ranking frequency (TOP2)

Similar to Stock and Watson (1999), we consider as a further criterion how often the squared forecast error for a particular IU metric $\left(\vartheta^\bullet_{i,\tau+\ell|\tau}\right)^2$ is among the two smallest of all alternative forecasting specifications. Then, the average $\mathrm{TOP2}^\bullet$ statistic is given by

$$\mathrm{TOP2}^\bullet = (1/((T-T_0+1)\times 18)) \sum_{\tau=T_0-\ell}^{T-\ell} \sum_{i=1}^{18} \mathrm{I}\left(\left(\vartheta^\bullet_{i,\tau+\ell|\tau}\right)^2 \leq \vartheta^2_{(2)i,\tau+\ell}\right), \tag{14}$$

where $\vartheta^2_{(2),i,\tau+\ell|\tau}$ is the 2nd-smallest squared forecast error obtained from six competing specifications including IU and one without IU.

4.1.2 Predictive accuracy in subsamples

The predictive content provided by individual IU quantifications might not be constant across time periods or economies. Lahiri and Sheng (2010) discuss situations in which alternative IU statistics are likely to feature distinct behavior. Moreover, the IU trajectories shown in Figures 3 and 4 highlight that time series and dispersion measures obtain particularly distinct IU estimates after 2008.

Thus, we first compare $\mathrm{RMSE}^\bullet$-statistics separately for either turbulent or tranquil subperiods. Such periods are distinguished by means of the standard deviation over the IU metrics in (2) to (8) at each forecasting step $\tau = T_0-\ell, ..., T-\ell$, denoted as $SD_{i,\tau+\ell|\tau}$. Forecast rankings are then determined by computing the average $\mathrm{RMSE}^\bullet$ separately for sample observations above and below $SD_{\tau+\ell|\tau}$, the median of $SD_{i,\tau+\ell|\tau}$ across economies. Second, the explanatory content of alternative IU statistics might depend on historical experiences of distinct economies with respect to the level of inflation. Therefore, we evaluate the predictive content of candidate IU measures separately for high- and low inflation economies.

4.2 Forecasting ex post uncertainty

Based on the alternative IU quantifications from (2) to (8), we obtain distinct forecasts $\widehat{\ln(e^2)}^{\bullet}_{\tau+\ell|\tau}$. Table 3 provides the corresponding RMSE statistics. The disparity statistic $\bar{\sigma}_{\tau+\ell|\tau}$ yields the most accurate forecasts. The predictive content of this IU metric is particularly visible for short anticipation horizons $\ell = 1$ to $\ell = 6$. We find considerable variation of the relative predictive performance of distinct measures of IU across forecast horizons ℓ. Since $\bar{\sigma}_{\tau+\ell|\tau}$ can be thought of as a combination of the IU statistics $\hat{\sigma}_{j,\tau+\ell|\tau}$, $j = 1, ..., J$, the success of this IU metric might result from the potential to average out individual disadvantages of these alternative IU statistics. The result also underscores the results in Becker and Clements (2008) for volatility forecasts. The rightmost column of the Tables 3 to 7 (labelled '$\delta_{2,k} = 0, \forall k$') documents the loss statistics that are obtained from the specification without IU. In almost all cases, forecasts are more accurate if ex ante IU metrics are employed than otherwise, i.e. for $\delta_{2,k} = 0, \forall k$.

Table 3: RMSE statistics

	$\hat{\sigma}_{\tau+\ell\|\tau}$	$\hat{h}_{G,\tau+\ell\|\tau}$	$\hat{h}_{SV,\tau+\ell\|\tau}$	$\hat{s}_{\tau+\ell\|\tau}$	$\bar{\sigma}_{\tau+\ell\|\tau}$	$\varsigma_{\tau+\ell\|\tau}$	$\delta_{2,k} = 0, \forall k$
ℓ	$B = 72$						
1	2.32	2.32	2.32	2.32	**2.28**	2.32	2.31
3	2.33	2.32	2.32	2.31	**2.25**	2.30	2.31
6	2.31	2.31	2.31	2.29	**2.26**	2.30	2.32
12	**2.43**	2.44	2.44	2.47	2.45	2.46	2.44
	$B = 108$						
1	2.25	2.25	2.25	2.24	2.24	2.24	**2.23**
3	2.39	2.38	2.39	2.37	**2.35**	2.38	2.38
6	2.31	2.30	2.30	2.30	**2.29**	2.31	2.32
12	2.23	2.23	**2.22**	2.26	2.27	2.26	2.22

NOTE: Cell entries document RMSE statistics, where the average is taken over $N = 18$ economies and forecast origins $\tau = T_0 - \ell, ..., T - \ell$, i.e. for the period between 1997M01 and 2014M12. The columns refer to the RMSE of forecasts based on the IU measures in (2) to (8). The estimation of these IU statistics is based on rolling samples of alternative length $B = \{72, 108\}$. The rightmost column ('$\delta_{2,k} = 0, \forall k$') shows the RMSE for predictions obtained by omitting IU from the forecasting specification, which serves as a benchmark device for the IU$^{\bullet}$-based forecasts. Each row contains results for one forecast horizon $\ell \in \{1, 3, 6, 12\}$. The smallest RMSE in each row is given in boldface.

Moreover, we find that the disagreement statistic $\hat{s}_{\tau+\ell|\tau}$ is less informative than the average uncertainty $\bar{\sigma}_{\tau+\ell|\tau}$. This is remarkable since disagreement is a particularly widely employed measure of IU (Lahiri and Sheng, 2010). Sample specific forecast rankings of the IU statistics are reported in Table 4. In most cases, the lead of $\bar{\sigma}_{\tau+\ell|\tau}$ over other IU measures is strongest at short to medium forecast horizons. However, the predictive accuracy of $\bar{\sigma}_{\tau+\ell|\tau}$ is not restricted to a particular subsample. In line with the full-sample results, the findings based on subsamples show that the disparity measure $\bar{\sigma}_{\tau+\ell|\tau}$ is more informative as a predictor variable for ex post IU than its time series based counterparts. Similarly, the distinction of turbulent and quiescent periods enables to draw clear-cut distinctions among the forecasts that are based on alternative IU metrics.

Table 4: RMSE statistics for subsamples ($B = 72$)

	$\hat{\sigma}_{\tau+\ell\vert\tau}$	$\hat{h}_{G,\tau+\ell\vert\tau}$	$\hat{h}_{SV,\tau+\ell\vert\tau}$	$\hat{s}_{\tau+\ell\vert\tau}$	$\bar{\sigma}_{\tau+\ell\vert\tau}$	$\varsigma_{\tau+\ell\vert\tau}$	$\delta_{2,k} = 0, \forall k$
ℓ	Turbulent periods						
1	2.29	2.29	2.30	2.29	**2.27**	2.28	2.27
3	2.29	2.28	2.28	2.27	**2.25**	2.28	2.27
6	2.29	2.31	2.30	2.25	**2.25**	2.25	2.30
12	**2.47**	2.48	2.48	2.51	2.51	2.50	2.47
	Quiescent periods						
1	2.35	2.34	2.34	2.34	**2.30**	2.35	2.35
3	2.37	2.36	2.36	2.35	**2.25**	2.32	2.36
6	2.32	2.32	2.32	2.34	**2.28**	2.34	2.34
12	2.40	2.41	2.41	2.44	**2.40**	2.42	2.41
	High inflation economies						
1	2.34	2.33	2.33	2.32	**2.29**	2.32	2.32
3	2.34	2.34	2.33	2.33	**2.26**	2.31	2.33
6	2.30	2.30	2.30	2.29	**2.25**	2.30	2.31
12	**2.39**	2.41	2.40	2.43	2.41	2.41	2.39
	Low inflation economies						
1	2.30	2.31	2.31	2.31	**2.27**	2.32	2.30
3	2.32	2.31	2.30	2.29	**2.24**	2.29	2.30
6	2.31	2.32	2.32	2.30	**2.28**	2.30	2.33
12	**2.48**	2.48	2.49	2.51	2.50	2.50	2.48

NOTE: Cell entries document RMSE statistics as described in Table 3 for subsamples, where each time instance $\tau = T_0 - \ell, ..., T - \ell$ is classified as a turbulent (quiescent) period if the standard deviation over all IU metrics from (2) to (8) lies above (below) its unconditional median. High- and low- inflation economies, respectively, are obtained by sorting the cross section into those economies which have an unconditional average inflation rate above or below the cross sectional median inflation. Results for a window length of $B = 108$ are qualitatively equivalent. Further details are reported in the notes to Table 3.

The findings obtained for the RMSE criterion are confirmed by the DA statistics documented in Table 5. In general, a DA in excess of 50% indicates that forecasts are valuable from an economic point of view. In this respect, all IU statistics deliver predictive content for horizons $\ell > 1$, with $\bar{\sigma}_{\tau+\ell|\tau}$ being the clearly most informative predictor also in terms of the DA criterion.

Table 5: Directional accuracy $\times 100$

	$\hat{\sigma}_{\tau+\ell\|\tau}$	$\hat{h}_{G,\tau+\ell\|\tau}$	$\hat{h}_{SV,\tau+1\|\tau}$	$\hat{s}_{\tau+\ell\|\tau}$	$\bar{\sigma}_{\tau+\ell\|\tau}$	$\varsigma_{\tau+\ell\|\tau}$	$\delta_{2,k}=0, \forall k$
ℓ	$B=72$						
1	51.02	49.72	49.26	49.63	**51.64**	49.63	48.76
3	51.79	51.41	53.30	53.95	**55.68**	53.86	51.57
6	**58.65**	58.62	57.60	57.82	58.24	57.57	58.30
12	53.57	54.50	54.07	54.53	**56.38**	54.89	52.58
	$B=108$						
1	48.36	47.55	48.32	48.08	**49.50**	47.89	45.47
3	52.98	53.11	51.73	53.92	**54.43**	53.45	51.48
6	56.90	57.09	57.18	**57.63**	57.25	56.13	54.47
12	55.56	54.66	54.93	55.89	**57.37**	57.04	52.91

NOTE: Cell entries show frequencies of directionally accurate forecasts in the sense of (13). Further details are reported in the notes to Table 3.

Moreover, the results in Table 6 show that $\bar{\sigma}_{\tau+\ell|\tau}$ is most often among the two most informative IU measures, in particular for anticipation horizons $\ell < 12$. We also compare the reported TOP2-frequencies to those obtained by drawing simulated forecasts under the assumption of equal predictive ability. As it can be seen in the Table, the IU-metric $\bar{\sigma}_{\tau+\ell|\tau}$ is most often in excess of the simulated critical values.

Table 6: TOP2 – percentages

	$\hat{\sigma}_{\tau+\ell\|\tau}$	$\hat{h}_{G,\tau+\ell\|\tau}$	$\hat{h}_{SV,\tau+\ell\|\tau}$	$\hat{s}_{\tau+\ell\|\tau}$	$\bar{\sigma}_{\tau+\ell\|\tau}$	$\varsigma_{\tau+\ell\|\tau}$	$\delta_{2,k}=0,\forall k$
ℓ	$B=72$						
1	30.40*	30.34*	29.07	21.88	**41.48***	22.90	23.92
3	21.91	23.83	25.68	30.86**	**40.93***	33.55***	23.24
6	25.40	26.20	26.39	32.87***	**35.12***	32.47***	21.54
12	27.62	27.50	27.69	30.65**	30.12	**31.45***	24.97
	$B=108$						
1	28.80	28.36	28.21	26.20	**35.56***	24.72	28.15
3	27.10	26.20	25.22	29.66	**40.80***	29.88	21.14
6	25.40	24.75	26.51	31.79**	**38.18***	30.62**	22.75
12	26.30	26.23	26.94	31.27***	**35.06***	33.36***	20.83

NOTE: Cell entries document frequencies by which IU measures are among the 'best 2'in terms of lowest RMSE statistics, as detailed in (14). The highest percentage in each row is in boldface. The significance of the TOP2-statistics is assessed by means of simulation where 10000 draws of squared forecast errors and corresponding TOP2-statistics are generated under the null hypothesis of equal predictive accuracy. Asterisks (*,**,***) denote the significance of the respective TOP2-numbers at the 10%, 5% and the 1%-level, respectively. Further details are reported in Table 3.

Table 7 documents in a further way how often inflation forecasts based on alternative IU measures differ significantly from each other. As for the results documented in Tables 3 to 6, differences in predictive accuracy are evaluated in terms of squared forecast errors, DA and the TOP2 criterion. The results show that consideration of the IU metric $\bar{\sigma}_{\tau+\ell|\tau}$ provides the highest number of cases where competing forecasts are significantly outperformed, especially at short- to medium forecast horizons.

To summarize, $\bar{\sigma}_{\tau+\ell|\tau}$ is the most informative predictor variable. The findings suggest that the accuracy of $\bar{\sigma}_{\tau+\ell|\tau}$ is not necessarily tied to the additional information provided by survey data, which is typically used to compute this type of IU statistics. Rather, this IU metric offers in many cases higher predictive content than its time series counterparts. As the results in this study suggest, the relative advantage of $\bar{\sigma}_{\tau+\ell|\tau}$ can also be found if alternative IU metrics are based on publicly available information with equal timing. Moreover, the documented findings are rather robust. The ranking of IU metrics is qualitatively identical for three distinct performance criteria. Similarly, $\bar{\sigma}_{\tau+\ell|\tau}$ remains the most informative IU statistic if the sample is split into subsets along the time series or the cross section dimension.

Table 7: Diebold and Mariano (1995) test results ($B = 72$)

	$\hat{\sigma}_{\tau+\ell\vert\tau}$	$\hat{h}_{G,\tau+\ell\vert\tau}$	$\hat{h}_{SV,\tau+\ell\vert\tau}$	$\hat{s}_{\tau+\ell\vert\tau}$	$\bar{\sigma}_{\tau+\ell\vert\tau}$	$\varsigma_{\tau+\ell\vert\tau}$	$\delta_{2,k} = 0, \forall k$
RMSE ℓ							
1	1	0	1	0	6	0	0
3	2	0	2	0	5	1	0
6	3	2	2	0	2	1	2
12	2	0	2	0	2	1	2
DA ℓ							
1	0	1	2	0	2	1	0
3	0	0	0	4	4	4	0
6	0	0	0	0	0	1	0
12	1	0	0	0	1	0	0
TOP2 ℓ							
1	3	3	4	1	6	0	0
3	1	1	1	4	6	4	0
6	3	1	1	4	5	5	0
12	2	0	0	3	3	3	0

NOTE: Cell entries document for a particular model, how many of the 6 respective alternative forecasts (i.e., 5 forecasts based on alternative IU measures in addition to the model without ex ante IU) are significantly outperformed with 5% significance. For $\ell > 1$, significance is evaluated by means of the adjusted DM statistic proposed by Harvey et al. (1997).

5 Conclusions

This study provides a comparative evaluation of alternative empirical measures of inflation uncertainty. In the related literature, various inflation uncertainty metrics have been distinguished. The first category of inflation uncertainty statistics consists of time series measures, the second is based on the heterogeneity of individual inflation forecasts. A descriptive analysis of several representatives of such ex ante measures of inflation uncertainty confirms this categorisation. Correlation statistics show that the relations among the members of both categories are stronger than across categories. We evaluate the alternative measures of inflation uncertainty by means of a performance comparison, where alternative metrics are considered as explanatory variables. The results show that, as a representative of the dispersion family, the average over individual uncertainties is the most preferable predictor. This result is in line with theoretical arguments of Lahiri and Sheng (2010) who compare the average individual uncertainty with the variation across point forecasts and find that, under certain assumptions, the latter may be regarded as an incomplete approximation of the former. The findings documented in our forecasting study are robust with respect to the choice of the forecasting evaluation criterion and also regarding distinct subsample choices. Furthermore, the trajectory of distinct IU statistics reveals that all uncertainty measures uniformly indicate low inflation uncertainty during the so-called Great Moderation and a subsequent uprise of inflation uncertainty around the year 2008.

Appendix

To approximate IU by means of the measures in (2) to (8), we employ a range of models for inflation forecasting. These models are listed in the following. Similar to Stock and Watson (2007), we extend the baseline AR in (1) with an output gap term, $\tilde{x}_t = x_t - \bar{x}_t$. This yields the backward looking Phillips curve, i.e.

$$\pi_{t+\ell} = \mu + \alpha_{21}(L)\pi_t + \alpha_{22}(L)\tilde{x}_t + \epsilon_{t+\ell}, \quad t = \tau - B + 1, ..., \tau, \tag{15}$$

with $\epsilon_{t+\ell} \sim (0, \sigma_\epsilon^2)$, $\alpha_{2k}(L) = \alpha_{2k,0} + \alpha_{2k,1}L + \cdots + \alpha_{2k,P}L^P,\ k = 1, 2$, and $P^{\max} = 12$. In (15), $\tilde{x}_t$ is estimated recursively based on observations dating in $t = \tau - B + 1, ..., \tau,\ B \in \{72, 108\}$, by means of the Hodrick-Prescott filter with the smoothing parameter set to 14400. An alternative model in the spirit of Cogley (2002) incorporates the so-called inflation gap, denoted $\tilde{\pi}_t = \pi_t - \bar{\pi}_t$, where $\bar{\pi}_t$ is the HP trend in π_t. The inflation gap model is given by

$$\pi_{t+\ell} = \mu + \alpha_{31}(L)\tilde{\pi}_t + \epsilon_{t+\ell}. \tag{16}$$

In this specification, $\pi_{t+\ell}$ is influenced by adjustment dynamics towards its long term trend, i.e. the relation in (16) is similar to an error correction relation.

A further model for the dynamics of inflation expectations is based on the concept of *adaptive expectations*. Following Branch (2004), adaptive inflation expectations denoted by π^e_{t+1}, are computed according to the recursion

$$\pi^e_{t+1} = w\pi_t + (1 - w)\pi^e_t. \tag{17}$$

The smoothing parameter w in (17) is determined by means of a grid search such that the in-sample MSE

$$\text{MSE} = (1/B) \sum_{t=\tau-B+1}^{\tau} (\pi_t - \pi_t^e)^2, \quad B \in \{72, 108\}, \tag{18}$$

is minimised at each step of the forecast recursion.

Branch (2004) also considers inflation forecasts that are obtained from vector autoregressive (VAR) models. We adopt this method by considering the VAR specification

$$y_{t+\ell} = \nu + A(L)y_t + v_t, \tag{19}$$

where $v_t \sim (0, \Sigma)$ and $A(L) = A_0 + A_1 L + \cdots A_P L^P$. As in (15), VAR models are estimated in rolling samples of size $B \in \{72, 108\}$. We specify a variety of alternative VAR models such that $y_t = (\pi_t, \tilde{x}_t, z_t)'$, i.e. inflation and the output gap are in all cases incorporated. Considering either one or two variables $z_t \in \{R_t, unemp_t, \Delta^2 oil_t, \Delta\overline{m}_t^{HP}\}$ obtains four three dimensional and six four dimensional VAR models.[6] In sum, we consider a total number of $J = 14$ inflation forecasting schemes.

References

Akqiray, V. (1989). Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business*, *62*, 55–80.

Ang, A., Bekaert, G., & Wei, M. (2007). Do Macro Variables, Asset Markets or Survey Forecast Inflation Better? *Journal of Monetary Economics*, *54*, 1163–1212.

Batchelor, R., & Dua, P. (1996). Empirical Measures of Inflation Uncertainty: A Cautionary Note. *Applied Economics*, *28*, 333–341.

Becker, R., & Clements, A. E. (2008). Are Combination Forecasts of S&P 500 Volatility Statistically Superior? *International Journal of Forecasting*, *24*, 122–133.

[6]As a quantification of adjustments in the long run trend of the money stock, the growth rate of the so-called 'core money', $\Delta\overline{m}_t^{HP}$, is obtained by means of applying the Hodrick-Prescott filter to the logarithm of a broad monetary aggregate.

Benati, L. (2008). The "Great Moderation" in the UK. *Journal of Money, Credit and Banking*, *40*, 121–147.

Blaskowitz, O., & Herwartz, H. (2014). Testing the Value of Directional Forecasts in the Presence of Serial Correlation. *International Journal of Forecasting*, *30*, 30–42.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, *77*, 623–685.

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

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Bomberger, W. A. (1996). Disagreement as a Measure of Uncertainty. *Journal of Money, Credit and Banking*, *28*, 381–392.

Branch, W. A. (2004). The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations. *The Economic Journal*, *114*, 592–621.

Brock, W. A., & Hommes, C. (1997). A Rational Route to Randomness. *Econometrica*, *65*, 1059–1095.

Canova, F. (2007). G-7 Inflation Forecasts: Random Walk, Phillips Curve or What Else? *Macroeconomic Dynamics*, *11*, 1–30.

Capistrán, C., & Ramos-Francia, M. (2010). Does Inflation Targeting Affect the Dispersion of Inflation Expectations? *Journal of Money, Credit and Banking*, *42*, 113–134.

Chevillon, G. (2005). Direct Multistep Estimation and Forecasting. *Journal of Economic Surveys*, *21*, 746–785.

Clements, M. P. (2014). Forecast uncertainty ex-ante and ex-post: U.S. inflation and output growth. *Journal of Business and Economic Statistics*, *32*, 206–216.

Clements, M. P., & Galvao, A. B. (2014). Measuring Macroeconomic Uncertainty: U.S. Inflation and Output Growth. *Available at SSRN 2436810*, .

Cogley, T. (2002). A Simple Adaptive Measure of Core Inflation. *Journal of Money, Credit and Banking*, *34*, 94–113.

D'Amico, S., & Orphanides, A. (2008). Uncertainty and Disagreement in Economic Forecasting. *Federal Reserve Board Finance and Economic Discussion Series Working Paper*, *56*.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

Diether, K. B., Malloy, C., & Scherbina, A. (2002). Differences of Opinion and the Cross Section of Stock Returns. *Journal of Finance*, *57*, 2113–2141.

Dovern, J., Fritsche, U., & Slacalek, J. (2012). Disagreement among Forecasters in G7 Countries. *Review of Economics and Statistics*, *94*, 1081–1096.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation. *Econometrica*, *50*, 987–1008.

Engle, R. F. (1983). Estimates of the Variance of U.S. Inflation Based upon the ARCH Model. *Journal of Money, Credit and Banking*, *15*, 286–301.

Friedman, M. (1977). Inflation and Unemployment. *Journal of Political Economy*, *85*, 451–472.

Giordani, P., & Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, *47*, 1037–1059.

Hamilton, J. (1985). Uncovering Financial Market Expectations of Inflation. *Journal of Political Economy*, *93*, 1224–1241.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281–291.

Herrera, A. M., & Pesavento, E. (2009). Oil Price Shocks, Systematic Monetary Policy, and the "Great Moderation". *Macroeconomic Dynamics*, *13*, 107–137.

Jurado, K., Ludvigson, S., & Ng, S. (2015). Measuring Uncertainty. *American Economic Review*, *105*, 1177–1216.

Knight, J., & Satchell, S. (2007). *Forecasting Volatility in the Financial Markets*. Butterworth-Heinemann.

Lahiri, K., & Liu, F. (2005). ARCH Models for Multi-Period Forecast Uncertainty: A Reality Check Using a Panel of Density Forecasts. In D. Terrell, & T. Fomby (Eds.), *Econometric Analysis of Financial and Economic Time Series Part A* (pp. 321–363). North Holland: Elsevier.

Lahiri, K., Peng, H., & Sheng, X. (2014). Measuring uncertainty of a combined forecast. Working Paper SUNY Albany.

Lahiri, K., & Sheng, X. (2010). Measuring Forecast Uncertainty by Disagreement: The Missing Link. *Journal of Applied Econometrics*, *25*, 514–538.

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.

Leitch, G., & Tanner, J. E. (1991). Economic Forecast Evaluation: Profits Versus the Conventional Error Measures. *American Economic Review*, *81*, 580–590.

Mankiw, N. G., & Reis, R. (2004). Disagreement about inflation expectations. *NBER Macroeconomics Annual*, *18*, 147–209.

McConnell, M. M., & Perez-Quiros, G. (2000). Output Fluctuations in the United States: What has changed since the early 1980s? *American Economic Review*, *90*, 1464–1476.

Mokinski, F., Sheng, X., & Yang, J. (2015). Measuring Disagreement in Qualitative Expectations. *Journal of Forecasting*, *34*, 405–426.

Okun, A. M. (1971). The Mirage of Steady Inflation. *Brookings Papers on Economic Activity*, *2*, 485–498.

Patton, A. J., & Sheppard, K. (2009). Original Combinations of Realised Volatility Estimators. *International Journal of Forecasting*, *25*, 218–238.

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

Rich, R., & Tracy, J. (2003). Modeling Uncertainty: Predictive Accuracy as a Proxy for Predictive Confidence. *Federal Reserve Bank of New York Staff Reports*, *253*.

Stock, J. H., & Watson, M. W. (1999). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In R. F. Engle, & H. White (Eds.), *Cointegration, Causality and Forcasting: A Festschrift in Honor of Clive W. J. Granger* (pp. 1–43). Oxford: Oxford University Press.

Stock, J. H., & Watson, M. W. (2007). Why has US Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, *39*, 3–33.

Stock, J. H., & Watson, M. W. (2008). Phillips Curve Inflation Forecasts. *NBER Working Paper*, *W14322*.

Swanson, N. R., & White, H. (1995). A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks. *Journal of Business and Economic Statistics*, *13*, 265–275.

Wallis, K. F. (2005). Combining Density and Interval Forecasts: A Modest Proposal. *Oxford Bulletin of Economics and Statistics*, *67*, 1389–1408.

Yu, J. (2011). Disagreement and Return Predictability of Stock Portfolios. *Journal of Financial Economics*, *99*, 162–183.

Zarnowitz, V., & Lambros, L. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, *95*, 591–621.