JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **29**: 736–757 (2014)
Published online 25 September 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2352

CONSTRUCTING OPTIMAL DENSITY FORECASTS FROM POINT FORECAST COMBINATIONS

WAGNER PIAZZA GAGLIANONE[a] AND LUIZ RENATO LIMA[b,c*]

[a] *Research Department, Banco Central do Brasil, Rio de Janeiro, Brazil*
[b] *Department of Economics, University of Tennessee at Knoxville, TN, USA*
[c] *Federal University of Paraiba, Brazil*

SUMMARY

Decision makers often observe point forecasts of the same variable computed, for instance, by commercial banks, IMF and the World Bank, but the econometric models used by such institutions are frequently unknown. This paper shows how to use the information available on point forecasts to compute optimal density forecasts. Our idea builds upon the combination of point forecasts under general loss functions and unknown forecast error distributions. We use real-time data to forecast the density of US inflation. The results indicate that the proposed method materially improves the real-time accuracy of density forecasts *vis-à-vis* those from the (unknown) individual econometric models. Copyright © 2013 John Wiley & Sons, Ltd.

Received 21 November 2011; Revised 1 July 2013

1. INTRODUCTION

Forecast combinations can be justified when the data-generating process is much more complex than any possible specification of individual models. In this case, each model is potentially misspecified and will certainly yield biased forecasts. Another important motivation for combining different forecasts arises when the information set of the individual models is private and therefore unknown to the decision maker. When this happens, the information set of the decision maker will only comprise individual forecasts, which suggests that an optimal forecast can be achieved by combining multiple predictions from different models. The success of forecast combination is widespread in such diverse areas as economics, finance, meteorology and political science, among others. Indeed, Clemen (1989, p. 559) points out that 'the results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy'. More recently, Stock and Watson (2001, 2004), Marcellino (2004) and Issler and Lima (2009) confirm Clemen's conclusion.

In a seminal paper, Granger and Ramanathan (1984) set out the foundations of optimal forecast combinations under symmetric and quadratic loss functions. They showed that under mean squared error (MSE) loss the optimal weights can be estimated through an ordinary least squares (OLS) regression of the target variable on a vector of forecasts plus an intercept to account for model bias. If the loss function differs from MSE, then the computation of optimal weights may require methods other than a simple OLS regression. In this paper, we derive optimal weights under general loss functions and unknown forecast error distributions. In this general framework, we are able to show that optimal weights can be easily identified through quantile regressions of the target variable on an intercept and a vector of individual point forecasts. This characterization of the optimal forecast combination as a quantile function is a generalization of the Granger and Ramanathan (1984) method and can be used to construct optimal density forecasts.

* Correspondence to: Luiz Renato Lima, Department of Economics, University of Tennessee at Knoxville, 527A Stokely Management Center, TN 37996-0550, USA.
E-mail: llima@utk.edu

Our research is related to the literature on combination of density forecasts initiated by Hall and Mitchell (2007), who derive optimal density forecasts based on the assumption of full knowledge of individual models and a fixed loss function. If the individual models are unknown, then one cannot estimate the individual densities, and therefore the approach suggested by Hall and Mitchell (2007) will not be feasible. This is exactly the case when an economic institution reports a point forecast but does not disclose the econometric model used to estimate it. This paper shows that we can use the information available on point forecasts to construct optimal density forecasts without requiring any knowledge of the (unknown) individual econometric models.

We applied the proposed forecast combination method to forecast the density of future US inflation. Such a forecast is affected by the fact that inflation volatility is not constant over time. As documented by Clark (2011), the volatility of inflation in the US remained extremely low during the 1988–2008 period due to the 'Great Moderation', but it has recently increased due to the increased volatility of energy prices. Thus, if the econometric model used to forecast inflation densities assumes constant variance, then such shifts in volatility will probably bias the density forecasts, making it too wide or too narrow. Another concern is that the distribution function of the data is probably unknown to the econometrician, but the current literature still places a parametric structure on the shape of the conditional distribution. If this parametric representation is misspecified, then density forecasts will probably be misleading. In this paper, we use the proposed approach to address these two issues jointly.

The evidence presented in this paper shows that the proposed method materially improves the real-time accuracy of density forecasts. More importantly, our empirical results indicate that the density forecast computed using our proposed method is outperformed neither by those constructed from the (unknown) individual econometric models nor by those obtained using combinations of densities or quantiles as suggested by Hall and Mitchell (2007) and Granger (1969, 1989), respectively. This empirical evidence is based on interval forecasts (coverage rates) and log predictive density scores.

This paper is organized as follows. Section 2 presents the forecast combination problem, discusses the econometric model and assumptions, and presents our results on optimal forecast combination. Section 3 discusses the estimation of optimal weights by using quantile regression (QR) and shows how to use the proposed method to construct density forecasts. Section 4 presents our empirical illustration and describes our real-time data. Section 5 presents the main results and section 6 concludes.

2. THE FORECAST COMBINATION PROBLEM

The decision maker (forecast user) is interested in forecasting at time t the future value of some stationary univariate time series $\{y_{t+h}\}_{h=1}^{\infty}$ on the basis of a k-vector of point forecasts of this variable $\hat{\mathbf{y}}_{t+h,t} = \left(\hat{y}^0_{t+h,t}, \hat{y}^1_{t+h,t}, \ldots, \hat{y}^{k-1}_{t+h,t}\right)$, which may include a constant. Note that each element of $\hat{\mathbf{y}}_{t+h,t}$ is determined *ex ante* (at time t) and is adapted to an expanding sequence of information sets $\mathcal{F}_t$. Hence $\hat{\mathbf{y}}_{t+h,t}$ (also denoted by $\hat{\mathbf{y}}_{t+h|t}$) is adapted to $\mathcal{F}_t$ whereas y_{t+h} is not, which rules out the uninteresting case where y_{t+h} is perfectly predictable. Thus the information set $\mathcal{F}_t$ comprises the k-vector of forecasts used to predict y_{t+h}. The decision maker seeks an aggregator that reduces the information in $\hat{\mathbf{y}}_{t+h,t} \in \mathbb{R}^k$ to a summary measure $C\left(\hat{\mathbf{y}}_{t+h,t}, \omega_i\right) \in \mathbb{R}$. This aggregator depends on the vector of weights $\omega_i \in \mathbb{R}^k$ whose identification, in turn, depends on both the loss function and the unknown forecast error distribution. We denote the conditional distribution of y_{t+h} given $\mathcal{F}_t$ as $F_{t+h,t}$, and the conditional density as $f_{t+h,t}$. Note that the parametric form of this conditional distribution (density) is unknown. A density forecast is therefore an estimate of $f_{t+h,t}$.

DOI: 10.1002/jae

Our premise is that, perhaps due to the presence of private information, the information set underlying individual forecasts is often unobserved to the decision maker. In this situation it is not feasible to pool the underlying information sets and construct a super model that nests each of the underlying forecasting models. Following Timmermann (2006), suppose that we are interested in forecasting the density of some variable, y_{t+h}, and that two point forecasts, $\hat{y}^1_{t+h,t}$ and $\hat{y}^2_{t+h,t}$, are available. Let the first point forecast be based on the variables x^1_t and x^2_t, i.e. $\hat{y}^1_{t+h,t} = g_1\left(x^1_t, x^2_t\right)$, whereas the second forecast is based on the variables x^3_t and x^4_t, i.e. $\hat{y}^2_{t+h,t} = g_2\left(x^3_t, x^4_t\right)$. If $\left\{x^1_t, x^2_t, x^3_t, x^4_t\right\}$ were available, it would be natural to estimate the quantile function of y_{t+h} based on all four variables, $Q_y(\tau) = g_3\left(x^1_t, x^2_t, x^3_t, x^4_t\right)$. The problem is that only the point forecasts $\hat{y}^1_{t+h,t}$ and $\hat{y}^2_{t+h,t}$ are observed by the decision maker (while the underlying variables are unobserved). The only option is then to estimate the quantiles by using only these forecasts, i.e. to elicit a quantile function of the type $q_y(\tau) = g_c\left(\hat{y}^1_{t+h,t}, \hat{y}^2_{t+h,t}\right)$. In other words, our premise is that the decision maker's information set, $\mathcal{F}_t$, comprises individual forecasts, $\mathcal{F}_t = \left\{\hat{y}^1_{t+h,t}, \hat{y}^2_{t+h,t}\right\}$, where $\mathcal{F}_t$ is often not the union of the information sets underlying the individual forecasts.

The conditional model with mean and variance dynamics is defined as

$$\begin{aligned} y_{t+h} &= \hat{\mathbf{y}}'_{t+h|t}\ \beta + \left(\hat{\mathbf{y}}'_{t+h|t}\gamma\right)\eta_{t+h} \\ \eta_{t+h}|\mathcal{F}_t &\sim\ F_{\eta,h}(0,1) \end{aligned} \tag{1}$$

where $F_{\eta,h}(0,1)$ is some distribution with mean zero and unit variance, which depends on h but does not depend on $\mathcal{F}_t$. $\hat{\mathbf{y}}_{t+h,t} \in \mathcal{F}_t$ is a $k \times 1$ vector of forecasts; β and γ are $k \times 1$ vectors of parameters which include the intercepts β_0 and γ_0. This class of DGPs is very broad and includes most common volatility processes such as ARCH and stochastic volatility. This location-scale model states that point forecasts $\hat{\mathbf{y}}_{t+h,t}$ affect both location and scale of the conditional distribution of y_{t+h}, which means that both conditional mean and conditional quantile functions will be affected by $\hat{\mathbf{y}}_{t+h,t}$.[1] An important thing to notice is that no parametric structure is placed on $F_{\eta,h}$.[2]

Following the literature (i.e. Granger, 1969; Granger and Newbold, 1986; Christoffersen and Diebold, 1997; Patton and Timmermann, 2007; Gaglianone and Lima, 2012), an optimal forecast combination $\hat{y}^i_{t+h,t}$ is obtained by minimizing the expected value of a general loss function L^i. In this paper we assume that such loss functions are defined according to Assumption 1 below.

Assumption 1 (loss function). The loss function L^i, where $i \in (0,1)$, is a homogeneous function solely of the forecast error $e_{t+h,t}$, i.e. $L^i = L^i(e_{t+h,t})$ and $L(ae) = g(a)L(e)$ for some positive function g.

Assumption 1 is exactly the same assumption L2 of Patton and Timmermann (2007). Although it rules out certain loss functions (e.g. those which also depend on the level of the predicted variable), it does include many common loss functions, such as MSE, MAE, lin-lin and many asymmetric quadratic losses. Index i is used to indicate that there exists a continuum amount of loss functions

[1] For a location model, only the conditional mean function will be affected by $\hat{\mathbf{y}}_{t+h,t}$.
[2] Elliott and Timmermann (2004) derived optimal weights based on the joint distribution $F\left(y_{t+h}, \hat{\mathbf{y}}'_{t+h,t}\right)$. They showed that it is the combination of asymmetry in both loss and data that is required for optimal weights to differ from the MSE weights. In this paper we show that some of the results in Elliott and Timmermann (2004) can also be obtained by only modeling the conditional distribution $F_{t+h,t}$.

DOI: 10.1002/jae

satisfying the above assumption and therefore a density forecast can be seen as a solution to the above decision maker's problem under various loss functions L^i, $i \in (0,1)$.[3]

Proposition 1. Under DGP(equation (1)) and a homogeneous loss function (Assumption 1), the optimal forecast combination will be

$$\hat{y}^{\,i}_{t+h,t} = \hat{\mathbf{y}}'_{t+h|t}\beta + (\hat{\mathbf{y}}'_{t+h|t}\gamma)\gamma^i_h = \omega_{i,0} + \omega_{i,1}\hat{y}^1_{t+h,t} + \ldots + \omega_{i,k-1}\hat{y}^{k-1}_{t+h,t} \tag{2}$$

where $\omega_{i,0} = \left(\beta_0 + \gamma_0\gamma^i_h\right); \omega_{i,1} = \left(\beta_1 + \gamma_1\gamma^i_h\right); \omega_{i,k-1} = \left(\beta_{k-1} + \gamma_{k-1}\gamma^i_h\right)$ and γ^i_h is a constant that depends only on the forecast horizon h, distribution $F_{\eta,h}(0,1)$ and the loss function L^i.

Proof. See Appendix.

Proposition 1 also nests some important special cases, as shown in the following corollaries.

Corollary 1. Under DGP (equation (1)) and the MSE loss function, the optimal forecast combination is

$$\hat{y}^{\,i}_{t+h,t} = E\left[y_{t+h}|\mathcal{F}_t\right] = \beta_0 + \beta_1\hat{y}^{\,1}_{t+h,t} + \ldots + \beta_{k-1}\hat{y}^{\,k-1}_{t+h,t}$$

where $E\left[y_{t+h}|\mathcal{F}_t\right]$ is the conditional mean of y_{t+h}. In this special case γ^i_h will correspond to $E\left[\eta_{t+h}|\mathcal{F}_t\right]$, which is zero by assumption. Therefore $\gamma^i_h = 0$ and optimal weights are fixed at $\omega_{i,j} = (\beta_j + \gamma_j.0) = \beta_j$, $j = 0, \ldots, k-1$.

Proof. See Appendix.

In the absence of scale effects in the conditional model (equation (1)), i.e. $\gamma_1 = \gamma_2 = \ldots = \gamma_{k-1} = 0$, the optimal forecast combination weights are identical to the MSE weights, and the intercept $\omega_{i,0}$ will still depend on the unknown distribution $F_{\eta,h}(0,1)$ and general loss function L^i. This latter result can be summarized in the following corollary.

Corollary 2. In the absence of scale effects, i.e. $\gamma_1 = \gamma_2 = \ldots = \gamma_{k-1} = 0$, the optimal forecast combination, under DGP(equation (1)), will be

$$\hat{y}^{\,i}_{t+h,t} = E\left[y_{t+h}|\mathcal{F}_t\right] + k^h_i = \omega_{i,0} + \beta_1\hat{y}^{\,1}_{t+h,t} + \ldots + \beta_{k-1}\hat{y}^{\,k-1}_{t+h,t} \tag{3}$$

where $\omega_{i,0} = \left(\beta_0 + \gamma_0\gamma^i_h\right)$ and $k^h_i = \gamma_0\gamma^i_h$.

Proof. See Appendix.

The above corollary shows that only the intercept depends on the shape of both the loss and distribution functions, while the forecast combination weights are equal to those obtained under MSE loss.

[3] This is not a strong restriction since we can always create many asymmetric losses by just adjusting the penalty received by negative and positive forecast errors.

DOI: 10.1002/jae

This same result was also obtained by Elliott and Timmermann (2004) using a joint distribution approach, but assuming a fixed forecast horizon and an elliptically symmetric distribution function. Our conditional model (1) delivers essentially the same result without assuming any functional form for $F_{\eta,h}(0,1)$ and without imposing a fixed forecast horizon.

3. ESTIMATION

Corollary 1 shows that, under MSE loss, the optimal combination corresponds to the conditional mean of y_{t+h}. The sample analog of the weights is of course the usual least squares estimator for the regression of y_{t+h} on a constant and the vector of forecasts, which was first proposed by Granger and Ramanathan (1984). Under general loss our Proposition 1 yields new results in the sense that the optimal weights will now differ from those obtained under MSE loss and the conditional model (1) offers a natural approach to estimating these optimal weights. Indeed, given the optimal forecast combination (2) and recalling that $F_{t+h,t}$ is the conditional distribution of y_{t+h}, we have the following result:

$$
\begin{aligned}
F_{t+h,t}\left(\hat{y}^{i}_{t+h,t}\right) &= \Pr\left(y_{t+h} < \hat{y}^{i}_{t+h,t}|\mathcal{F}_t\right) \\
&= \Pr\begin{pmatrix} \hat{\mathbf{y}}'_{t+h|t}\beta + \left(\hat{\mathbf{y}}'_{t+h|t}\gamma\right)\eta_{t+h} < \\ < \hat{\mathbf{y}}'_{t+h|t}\beta + \left(\hat{\mathbf{y}}'_{t+h|t}\gamma\right)\gamma^{i}_{h}|\mathcal{F}_t \end{pmatrix} \\
&= \Pr\left(\eta_{t+h} < \gamma^{i}_{h}|\mathcal{F}_t\right) = F_{\eta,h}\left(\gamma^{i}_{h}\right) \\
&= \tau_i \in (0,1)
\end{aligned} \tag{4}
$$

Thus it follows that $\hat{y}^{i}_{t+h,t} = F^{-1}_{t+h,t}(\tau_i)$, which means that the optimal forecast combination $\hat{y}^{i}_{t+h,t}$ coincides with the conditional quantile function of y_{t+h} at level τ_i, i.e.

$$
\begin{gathered}
\hat{y}^{\,i}_{t+h,t} = Q_{y_{t+h}}(\tau_i|\mathcal{F}_t) = \omega_{i,0} + \omega_{i,1}\hat{y}^{\,1}_{t+h,t} + \ldots + \omega_{i,k-1}\hat{y}^{\,k-1}_{t+h,t} \\
\text{for some } \tau_i \in (0,1)
\end{gathered} \tag{5}
$$

where $Q_{y_{t+h}}(\tau_i|\mathcal{F}_t)$ is the conditional quantile of y_{t+h} at level τ_i and the weights are estimated as in equation (2) with γ^{i}_{h} equal to $F^{-1}_{\eta,h}(\tau_i)$.[4] Thus the optimal weights under general loss L^i and unknown distribution function $F_{\eta,h}(0,1)$ can be obtained through a QR of y_{t+h} on a constant and the vector of forecasts.[5]

Remark 1. Equation (5) generalizes the idea of Granger and Ramanathan (1984), who employed OLS to estimate MSE weights. Under general loss L^i and unknown distribution function $F_{\eta,h}(0,1)$, the optimal weights can be estimated using the quantile regression method proposed by Koenker and Basset (1978).

As an example, if $\tau_i = 0.5$ then $\hat{y}^{\,i}_{t+h,t} = Q_{y_{t+h}}(0.5|\mathcal{F}_t)$ is the conditional median of y_{t+h}, which is an optimal forecast under the mean absolute error (MAE) loss function. In this case, the optimal weights will be given by $\omega_{i,j} = \left(\beta_j + \gamma_j F^{-1}_{\eta,h}(0.5)\right)$ $j = 0, \ldots, k-1$, where $F^{-1}_{\eta,h}(0.5)$ is the median of η_{t+h}. If a loss function penalizes positive errors more than negative ones, then the optimal forecast combination will correspond to $Q_{y_{t+h}}(\tau_i|\mathcal{F}_t)$ with $\tau_i > 0.5$. The more one penalizes the positive errors, the bigger τ_i

[4] From the third line in equation (4) one can see that $F_{\eta,h}\left(\gamma^{i}_{h}\right) = \tau_i$, which implies that $\gamma^{i}_{h} = F^{-1}_{\eta,h}(\tau_i)$.

[5] The higher the degree of overlap in the information sets used to produce the underlying point forecasts, the less useful a combination of forecasts is likely to be.

DOI: 10.1002/jae

will be and, in this case, the optimal weights will be given by $\omega_{i,j} = \left(\beta_j + \gamma_j F^{-1}_{\eta,h}(\tau_i)\right)$, $\tau_i > 0.5$. On the other hand, if the decision maker's loss function penalizes negative forecast errors more than positive ones, then the optimal forecast combination will be given by $Q_{y_{t+h}}(\tau_i|\mathcal{F}_t)$ with $\tau_i < 0.5$ and optimal weights would be computed accordingly. Considering all possible loss functions (i.e. all possible quantiles $\tau_i \in (0,1)$) we obtain the following density forecast (gray area) shown in Figure 1.[6]

The main idea is that, following Elliott and Timmermann (2008), two forecast users with different loss functions will want different quantiles of the distribution, so a single number could never give them both the optimal forecasts. However, it would be sufficient to provide the entire distribution (density forecast) because this has all the quantile information they are looking for. Of course, under MSE loss, all forecast users agree that the mean of the predictive density is the most important information. A nice aspect of this methodology is that, since we are using QR to estimate optimal weights, the functional form of $F_{\eta,h}$ is left unspecified.

The forecasting literature has also suggested approaches that combine probability density forecasts.[7] Stone (1961) considered the following linear combination:

$$f^c_{t+h,\,t}\left(y_{t+h}\right) = \sum_{j=1}^{k} \omega_{t+h,\,t,\,j} f_{t+h,\,t,\,j}\left(y_{t+h}\right) \tag{6}$$

where $f_{t+h,t,j}$ is the conditional density from the jth model and $\omega_{t+h,t,j}$ are weights. Hall and Mitchell (2007) proposed combining predictive probability densities by finding weights $\omega_{t+h,t,j}$ that maximize the average log predictive score function as follows:

$$w^* = \arg\max_{w} \frac{1}{T} \sum_{t=1}^{T} \ln\left(f^c_{t+h,t}\left(y_{t+h}\right)\right) \tag{7}$$

They show that when weights are chosen as in equation (7) then the combined density is optimal, in the sense that it minimizes the following loss function:

$$\overline{KLIC} = \frac{1}{T} \sum_{t=1}^{T} \left[\ln f_{t+h,\,t}\left(y_{t+h}\right) - \ln f^c_{t+h,\,t}\left(y_{t+h}\right)\right] \tag{8}$$

Thus the above method finds a set of weights on the n-dimensional simplex maximizing the log score of the linear combination or, equivalently, minimizing the $\overline{\text{KLIC}}$ distance of the linear combination to the unknown data generation process.[8]

Another related approach was suggested by Granger (1969, 1989). In other words, if we let $q^\tau_{t+h,t,1}, \ldots, q^\tau_{t+h,t,k}$ be a set of τth quantile forecasts from k models, then a common practice in the forecasting literature is to consider the following quantile forecast combination:

$$C^\tau_{t+h,\,t} = \sum_{j=1}^{k} \omega_j(\tau) q^\tau_{t+h,\,t,\,j} \tag{9}$$

where ω_j are weights that can be estimated using the method suggested by Granger (1969, 1989). While equation (9) may work in practice, it seems a little hard to justify it from a distributional point

[6] The dark lines represent the out-of-sample empirical quantiles $\tau = \{0.25; 0.50; 0.75\}$.

[7] We do not discuss here the methods proposed by Andrade *et al*. (2012) because it is implementable only when individual densities are available.

[8] What may be problematic in this method is that, due to the insufficient capacity of the pool of combined models to provide different relative forecasting properties at different times, the optimal weights might be poorly identified in many applications. We sincerely thank an anonymous referee for pointing this out.

DOI: 10.1002/jae

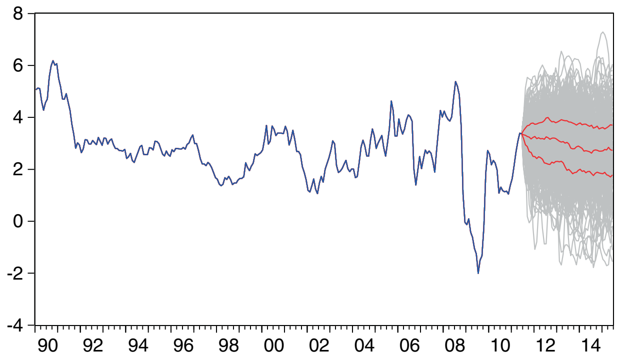

Figure 1. Inflation rate density forecast example

of view. Suppose we have two models and each one generates two different densities. If we consider the mixture of those densities, the τth percentile of the combined distribution will not in general correspond to the weighted average of the τth percentiles of the individual distributions. In short, the quantile of the sum may not be equal to the sum of the quantiles.

Finally, the feasibility of the above methods depend on full knowledge of econometric models used to generate the individual densities (or quantiles). Our approach, on the other hand, relies only on individual point forecasts, which is particularly important in situations where the decision maker wants to construct optimal density forecasts without making assumptions on the parametric specification of the individual (unknown) models.

4. FORECASTING INFLATION DENSITIES

The main purpose of this section is to provide empirical evidence for the theoretical results previously derived. To this end, we will consider the forecast of inflation densities because it is crucial to economic decision making as a measure of inflation uncertainty. Nominal interest rates tend to be higher when uncertainty about future inflation is higher and, therefore, investment decisions in both money and capital markets are obviously affected. Judson and Orphanides (1999) analyze the effects of the volatility of inflation on economic growth and document evidence in favor of the hypothesis that uncertainty on future inflation hurts economic growth. Uncertainty about inflation can also affect fiscal policy planning in the sense that it increases the unpredictability of future fiscal revenue. As conjectured by Milton Friedman (1977), an increase in inflation uncertainty reduces economic efficiency and possibly output growth. In order to avoid the negative effects of inflation uncertainty, many central banks, such as the Bank of Canada, Bank of England, Norges Bank, Central Bank of Brazil and Sveriges Riksbank, are now publishing fan charts that provide entire forecast distributions for inflation. These fan charts can be used to forecast the probability that future inflation will fall within an interval pre-specified by the central bank.

Forecasts of inflation density in the USA are affected by two major problems. First, as documented by Clark (2011), the volatility of inflation in the USA decreased during the 1988–2008 period due to the event known as the 'Great Moderation'. However, more recently, increased volatility of energy prices has caused the volatility of total inflation to rise sharply. If the econometric model used to forecast inflation densities assumes constant variance, then such shifts in volatility will probably bias the density forecasts, making it too wide or too narrow. Recent research, i.e. Carriero *et al*. (2012a,2012b), Clark (2011) and Jore *et al*. (2010), has shown that density forecasts are improved when one allows for the presence of time-varying variance. A second concern is that the distribution function of the future inflation is probably unknown to

DOI: 10.1002/jae

the econometrician. The current literature, although allowing for a time-varying variance, still places a parametric structure on it. If this parametric representation is misspecified, then density forecasts will probably be misleading.

In this section we address these two concerns by using the optimal forecast combination method proposed previously. Recall that the proposed method is based on the location-scale model (1), which allows for the covariates to affect both the location and the scale of the distribution function and therefore addresses the first concern. Estimation of the model uses QR methods which do not require knowledge of $F_{\eta,h}$ and therefore address the second concern. Moreover, the combination device used in our approach contributes to minimizing the uncertainty about the correct specification of the conditional quantile function in the same way that the forecast combination method of Granger and Ramanathan (1984) was used to minimize the uncertainty about the correct specification of the conditional mean function. Indeed, as stressed by Stock and Watson (2001, 2004), individual forecasting models may be subject to misspecification bias of unknown form. Therefore combining forecasts across different models can be viewed as a way to make the forecast more robust against such a misspecification bias (Timmermann, 2006).

4.1. The Econometric Models

In this section, we will assume that there are fictitious economic institutions that use different econometric models to make forecasts. We will pretend that the decision maker only observes the point forecasts from each institution, here represented by the conditional mean of each model. Next, we will assume that there is another agent that has full information about all individual models and therefore he or she will be able to combine individual density forecasts. Thus the contribution of each individual model introduced in this section will be twofold.[9] First, they will be employed to forecast the conditional mean, which is used as covariate in equation (5). Second, the models will be used to forecast the densities of inflation that will be further combined by using the approach suggested by Hall and Mitchell (2007).

Following Faust and Wright (2012), all models considered in this paper are focused on inflation 'gaps' (instead of directly modeling inflation rates) defined here as $g_t = \pi_t - \tau_t$, where g_t is the inflation gap (treated as stationary), π_t is the inflation rate and τ_t is the respective inflation trend. According to the referred authors, the idea is to forecast the inflation gap around some slow-varying local mean (i.e. low-frequency component), which has found to be quite a successful approach (see Faust and Wright, 2012, p. 10, for further details).[10] The inflation trend, which is often interpreted as representing an agent's perceptions of the Fed's long-run inflation target, is here proxied by a moving average process based on the previous 4 years of the observed real-time inflation rate.[11] Along the out-of-sample forecast horizon h the inflation trend τ_t is assumed to follow a random walk (i.e. $E[\tau_{T+h}|\mathcal{F}_T] = \tau_T$ for all h).

Thus we first estimate conditional models to the inflation gap and then use them to make direct forecasts for several forecast horizons h (see Marcellino *et al*., 2006), which are further added to the respective inflation trend forecast in order to obtain the inflation forecasts. The conditional models for the inflation gap can be cast in the following first general setup:

---

[9] The term 'model' is here used in a broad sense that includes forecasting methods.

[10] We sincerely thank an anonymous referee for this suggestion.

[11] Alternative trends could also be constructed from survey-based long-run expectations (e.g. SPF or Blue Chip), although exhibiting short sample sizes in many cases.

DOI: 10.1002/jae

$$
\begin{aligned}
y_{t+h} &= g_{t+h} = \pi_{t+h} - \tau_{t+h} = X'_{t+h,t}\alpha + \eta_{t+h} \\
\eta_{t+h} &\sim \mathrm{Normal}\left(0, \sigma^2_{t+h}\right)
\end{aligned}
\tag{10}
$$

where π_{t+h} is the inflation rate at some future time $t+h$, τ_{t+h} is respective inflation trend and $X_{t+h,t}$ is a vector of economic indicators that are known at time t. Note that this approach allows for heteroskedastic dynamics, in which the conditional variance is assumed to follow a standard Gaussian GARCH(1,1) process.[12] Model (10) admits a great variety of specifications and hence we will consider the following ones: [13]

- *Model 1*. AR(1) with fixed $\rho = 0.46$: The first model is a simple AR(1) with fixed autoregressive coefficient ρ. We assume that the inflation gap follows an AR(1) with a fixed slope coefficient set to 0.46 (this follows model 7 of Faust and Wright, 2012). This is the model (10) with $X'_{t+h,t} = g_t$ and $\alpha = 0.46$.
- *Model 2*. AR(2): This is the model (10) with $X'_{t+h,t} = (1, g_t, g_{t-1})$ and $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$.
- *Model 3*. RW-AO: This is the variant of the pure random walk (RW) model considered by Atkeson and Ohanian (2001), which is model (10) with $X'_{t+h,t} = \frac{1}{4}\sum_{j=1}^{4} g_{t-j+1}$ and $\alpha = 1$.
- *Model 4*. PC (backward looking): This model is the Phillips curve (PC), which has a long tradition in forecasting inflation. The Phillips curve has been exhaustively used to make point forecasts of future inflation (for a comprehensive survey, see Stock and Watson, 2008) This is model (10) but with $X'_{t+h,t} = (1, g_t, u_t)$ and $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$, where u_t is the unemployment rate.
- *Model 5*. PC (backward looking, extended): Same as model 4, but with $X'_{t+h,t} = (1, g_t, g_{t-1}, u_t, u_{t-1})$ and $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)'$.

In models 4 and 5 only 'adaptive expectations' for the inflation gap and the unemployment rate matter to forecasting the future values of inflation gaps. We also consider three additional models taking into account inflation expectations:

- *Model 6*. PC-hybrid (backward and forward looking): This is the model (10) but with $X'_{t+h,t} = (1, g_{t+h,t}, g_t, u_t)$ and $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)'$, where $g_{t+h,t}$ is the market expectation of the inflation rate (i.e. $\pi_{t+h,t}$ defined as the mean forecast from the Survey of Professional Forecasters (SPF), published by the FED of Philadelphia) subtracted by the respective forecast of the inflation trend ($\tau_{t+h,t}$), i.e. $g_{t+h,t} = \pi_{t+h,t} - \tau_{t+h,t}$.
- *Model 7*. PC-hybrid (backward and forward looking, extended): This is the model (10) but with $X'_{t+h,t} = (1, g_{t+h,t}, g_t, u_t, \Delta u_t)$ and $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)'$.
- *Model 8*. Survey-based expectations (SPF): This is the model (10) but with $X'_{t+h,t} = g_{t+h,t}$ and $\alpha = 1$.

As mentioned, models 1–8 are estimated by using a Gaussian-GARCH(1,1) approach. Nonetheless, in order also to allow for asymmetric dynamics in the inflation gap, we re-estimate these previous models by using the so-called two-piece normal distribution approach employed by the Bank of England. According to Britton *et al*. (1998), the probability density function for the so-called two-piece-normal-distribution is given by AS, in which $A = \frac{2}{\left(\left(1/\sqrt{1-\gamma}\right)+\left(1/\sqrt{1+\gamma}\right)\right)}$ and $S = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[\frac{-1}{2\sigma^2}\left\{(x-\mu)^2 + \gamma\left(\frac{x-\mu}{|x-\mu|}\right)(x-\mu)^2\right\}\right]}$.

[12] For the estimation of the GARCH(1,1) models based on real-time data, we use the variance targeting technique in order to increase estimation stability. This approach restricts the constant term of the conditional variance equation to a function of the GARCH parameters and the unconditional variance of the residuals.

[13] Section 4.2 presents a complete description of the data used to estimate each individual model.

DOI: 10.1002/jae

In this paper, we construct conditional density forecasts based on the Bank of England's approach (models 9–16) by assuming that: (i) the conditional mean μ is given by those of models 1–8 (respectively); (ii) the conditional variance σ^2 is determined by GARCH(1,1) estimates of models 1–8 and; (iii) the conditional skew (γ), which measures the degree of asymmetry of the conditional density, is given by the sample skew (skw), which is based on the (real-time) GDP price index inflation rate of the previous ten years. It is also normalized as $\gamma = \frac{\text{skw}}{1+|\text{skw}|}$, which guarantees that $\gamma \in [-1; 1]$. Note that if the conditional skew is set to zero, then models 9–16 simplify to models 1–8.

Thus, given the empirical evidence that the volatility of inflation is not constant over time, models 1–8 exhibit a time-varying volatility with symmetric distribution, whereas models 9–16 have a time-varying volatility with an asymmetric distribution. Furthermore, all models assume a parametric form for the error distribution, which may differ from the true one, and this also affects the accuracy of density forecasts. We also estimate each model by using QR,[14] which is a semi-parametric approach since it assumes that the parametric form of the error distribution is unknown.

We are not claiming that the above suite of models is the best one and we admit that more models could be added to it. For example, we could specify the Phillips curve models using economic leading indicators other than the unemployment rate. Although we think that this extension would be valuable, the above list also seems to be a reasonable approximation to the spectrum of models used by commercial banks and other economic institutions. In what follows, each of the previous models will be used to estimate point forecasts $\hat{\mathbf{y}}_{t+h,t}$, conditional quantiles, $q^{\tau}_{t+h,t}$, for $\tau \in (0,1)$ and densities $f_{t+h,t}$.

We now turn to the forecast combination schemes. In this sense, we consider the sets of densities, $S^d_1 = \{f_{t+h,t,j}\}_{j=1,\dots,8}$, $S^d_2 = \{f_{t+h,t,j}\}_{j=9,\dots,16}$, $S^d_3 = \{f_{t+h,t,j}\}_{j=1,\dots,16}$, a set of point forecasts $S^{pf}_4 = \left\{\hat{y}^j_{t+h,t}\right\}_{j=1,\dots,8}$, and a set of quantiles $S^q_5 = \left\{q^{\tau}_{t+h,t,j}\right\}_{j=1,\dots,8}$ for $\tau \in (0,1)$.

For the combination method (equation (6)), we follow Kascha and Ravazzolo (2010) and consider three weighing schemes: equal weights, MSE weights and recursive log score weights, all applied to the sets of densities S^d_1, S^d_2 and S^d_3. Thus models 17, 18 and 19 are the combined densities obtained using equal weights, based on sets S^d_1, S^d_2 and S^d_3, respectively. In the same way, models 20, 21 and 22 are combined densities using MSE weights, and models 23, 24 and 25 are combined densities using recursive log score weights.

The combination method of Elliott and Timmermann (2004) (equation (3)) is based on S^{pf}_4 and produces model 26. The optimal combination approach (equation (5)) is based on the same set S^{pf}_4 and generates our model 27. The quantile combination approach proposed by Granger (1969, 1989) is labeled model 28. For models 26, 27 and 28, following Koenker (2005), given a family of estimated conditional quantile functions, the conditional density of y_{t+h} can be estimated by the formula

$$\hat{f}_{t+h,t} = \frac{(\tau_i - \tau_{i-1})}{\hat{Q}_{y_{t+h}}(\tau_i|\mathcal{F}_t) - \hat{Q}_{y_{t+h}}(\tau_{i-1}|\mathcal{F}_t)}$$

where $\hat{Q}_{y_{t+h}}(\tau_i|\mathcal{F}_t)$ and $\hat{Q}_{y_{t+h}}(\tau_{i-1}|\mathcal{F}_t)$ are two adjacent estimated conditional quantiles of y_{t+h}. The conditional density $\hat{f}_{t+h,t}$ can also be estimated (for instance) by the Epanechnikov kernel, which is a weighting function that determines the shape of the bumps. We prefer the latter because it generates smooth densities, especially when the time series sample size is short, which is the case in this empirical application of inflation forecasting. Tables I and II summarize the individual forecasting models as well as the combination approaches. In the next section, we describe the database as well as the forecasting schemes employed in the empirical exercise.

[14] In order to construct the density forecast combination scheme proposed by Granger (1969, 1989).

DOI: 10.1002/jae

Table I. Conditional models with Gaussian GARCH(1,1) estimation

Model	Label	Covariate vector $X'_{t+h,t}$
1	AR(1) with fixed $\rho=\alpha=0.46$	g_t
2	AR(2)	$(1, g_t, g_{t-1})$
3	RW-AO, $\alpha=1$	$\frac{1}{4}\sum_{j=1}^{4} g_{t-j+1}$
4	PC (backward looking)	$(1, g_t, u_t)$
5	PC (backward looking, extended)	$(1, g_t, g_{t-1}, u_t, u_{t-1})$
6	PC-hybrid (back-forward looking)	$(1, g_{t+h,t}, g_t, u_t)$
7	PC-hybrid (back-forward looking, extended)	$(1, g_{t+h,t}, g_t, u_t, \Delta u_t)$
8	Survey-based expectations (SPF), $\alpha=1$	$g_{t+h,t}$

Note: Each model $i=9$ to 16 has the same (respective) covariate vector of model $i-8$ but is estimated using Britton *et al.*'s (1998) approach and thus allows for time-varying conditional variance with asymmetric dynamics.

4.2. Data

Inflation is measured with the GDP or GNP deflator, depending on data vintage, and the inflation rate is defined as the annualized quarterly rate, defined as $400\times\log(P_{t+h}/P_{t+h-1})$, $h=1,2,3,4,8$ and 12, in which P_t is the (real-time) GDP (or GNP) price index. All other variables are also log-transformed. We collect our data from the Federal Reserve Bank of Philadelphia's Real Time Data Set for Macroeconomists (RTDSM). In Figure 2 we display the realized inflation π_t, the inflation trend τ_t, and the inflation gap g_t. Figure 2 suggests that inflation gap is very volatile from 1971 to 1988, pretty stable during the great moderation (1988–2008) and presents an increase in volatility more recently with the mortgage crisis. These facts point out that the volatility of g_t is not constant over time and therefore models that fail to account for time-varying volatility will not present a good forecasting performance.

We also use data on quarterly market forecasts of annualized inflation rates, measured as the respective mean forecasts from the Survey of Professional Forecasters (SPF), published by the Federal Reserve Bank of Philadelphia. Since the SPF dataset begins in 1968:Q4, for model estimation purposes we extended the time series of GDP price index expectations from 1968:Q3 back to 1960:Q1 by using exponential smoothing.[15] Since the SPF does not provide expectations for $h=8$ and 12 quarters ahead, we first set the inflation expectation for $h=12$ equal to the CPI5 year, and use an exponential smoothing to backcast past missing values. Then, we use a 'term-structure' linear interpolation to generate expectations for $h=8$ based on observed figures for $h=4$ and $h=12$.

The starting point of the estimating sample is always 1961:Q1. In order to use the sample 1961:Q1–1984:Q4, we adopt vintage 1985:Q1 (with information until 1984:Q4). The only exception is vintage 1996:Q1, due to data unavailability, in which information for GDP price index in 1995:Q4 is obtained from the next vintage. We conduct a 'pseudo out-of-sample' exercise in which forecasts are generated both by a recursive scheme (i.e. expanding sample size) as well as by a rolling (20 years) sample scheme. In the former, the individual models are initially estimated by using a sample that starts at 1961:Q1 and ends at 1984:Q4, but it is expanded as we go into the out-of-sample period. In the latter, we keep the estimating sample size constant at 80 observations (20 years) and then we discard and add the oldest and newest observations, respectively, as we go into the out-of-sample period. The full forecast evaluation runs from 1985:Q1 to 2012:Q1. For each forecast starting at the origin t = 1985:Q1, we use the real-time data vintage t, which

[15] Following Clark (2011), the exponentially smoothed series for the expected inflation rate $\pi^*_{t+1,t}$ is constructed as follows: (i) initialize the filter with the average inflation rate (π_t) of 1953:Q1–1959:Q3. The average becomes the exponentially smoothed estimate for period 1959:Q4; (ii) use exponential smoothing formula $\pi^*_{t+1,t} = \alpha\pi_t + (1-\alpha)\pi^*_{t,t-1}$ with a (calibrated) smoothing parameter $\alpha=0.30$ to estimate the trend inflation expectation for 1960:Q1 based on *t-1*; (iii) define the remaining values for next period t as the exponentially smoothed trend estimated with data through t-1. For longer forecast horizons ($h>1$), we assume that $\pi^*_{t+h,t} = \alpha\pi^*_{t+h-1,t} + (1-\alpha)\pi^*_{t+h-1,t-1}$.

DOI: 10.1002/jae

Table II. Conditional models from combination approaches

Model	Weight	Set of base models: density f, point forecast $\hat{y}$ or quantile q
17	Equal	$S_1^d = \{f_{t+h,t,j}\}_{j=1,\ldots,8}$
18	Equal	$S_2^d = \{f_{t+h,t,j}\}_{j=9,\ldots,16}$
19	Equal	$S_3^d = \{f_{t+h,t,j}\}_{j=1,\ldots,16}$
20	MSE	$S_1^d = \{f_{t+h,t,j}\}_{j=1,\ldots,8}$
21	MSE	$S_2^d = \{f_{t+h,t,j}\}_{j=9,\ldots,16}$
22	MSE	$S_3^d = \{f_{t+h,t,j}\}_{j=1,\ldots,16}$
23	Log score	$S_1^d = \{f_{t+h,t,j}\}_{j=1,\ldots,8}$
24	Log score	$S_2^d = \{f_{t+h,t,j}\}_{j=9,\ldots,16}$
25	Log score	$S_3^d = \{f_{t+h,t,j}\}_{j=1,\ldots,16}$
26	E&T (2004)	$S_4^{\mathrm{pf}} = \left\{\hat{y}_{t+h,t}^j\right\}_{j=1,\ldots,8}$
27	Optimal QR	$S_4^{\mathrm{pf}} = \left\{\hat{y}_{t+h,t}^j\right\}_{j=1,\ldots,8}$
28	Granger	$S_5^q = \left\{q_{t+h,t,j}\right\}_{j=1,\ldots,8}$

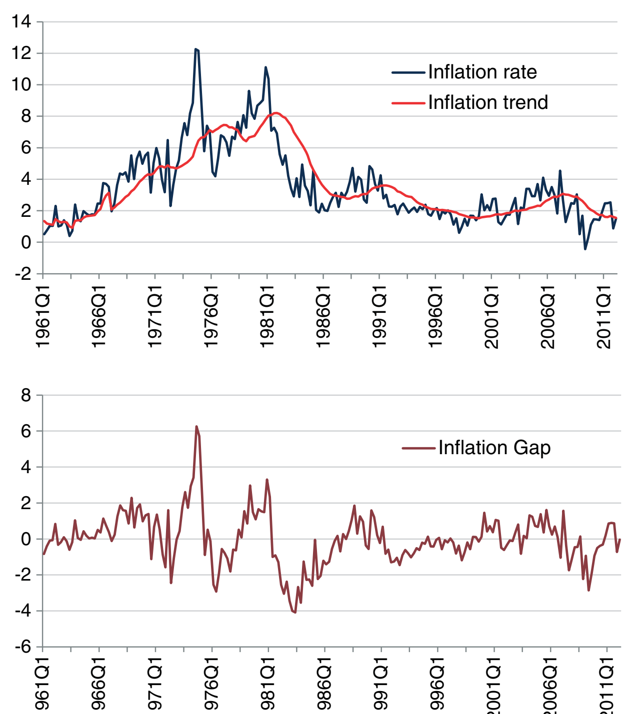

Figure 2. Inflation rate (% p.a.), inflation trend and inflation gap. Note: vintage 2012:Q2 is used to compute the inflation rate, trend and gap

usually contains information up to *t-1*, to estimate the models and construct forecasts for periods t and beyond. Finally, in order to estimate the weights for the forecast combinations in models 17–28, a training sample of TS = 60 observations is considered. This way, for any h, the training sample of the dependent variable y_{t+h} will range from 1985:Q1 $-$ TS + 1 to 1985:Q1 and from 1985:Q1 $-$ TS + 1 $-$ h to 1985:Q*1-h* for the individual model forecasts.

5. RESULTS

The out-of-sample forecasting exercise is made by using both a recursive and rolling scheme. We conduct forecast evaluations based on the entire density, which includes: (i) coverage rates; (ii) log predictive density score, which allows one to rank the models; and (iii) the Amisano and Giacomini (2007) test, which compares the log score distance between two models. Our forecast evaluation is conducted using real-time data. According to Clark (2011): 'evaluating the accuracy of real-time forecasts requires a difficult decision on what to take as the actual data in calculating forecast errors'. Thus we adopt the second available estimates of GDP/GNP deflator as actuals in evaluating forecast accuracy. For instance, for the case in which h-step-ahead forecasts are made for period $t+h$ with vintage t data ending at period t-1, the second available estimate will be taken from the vintage $t+h+2$ dataset.

A natural starting point for forecast density evaluation is interval forecasts, i.e. coverage rates. Recent studies such as Giordani and Villani (2010) and Clark (2011) have used interval forecasts as a measure of the calibration of macroeconomic density forecasts (see also Mitchell and Wallis, 2010). Table III reports the frequency with which actual real-time outcomes for inflation falls inside the 90% interval. Accurate intervals should result in frequencies of about 90%. A frequency of more (less) than 90% means that, on average over a given sample, the density is too wide (narrow). The table includes p-values for the null of correct coverage of 90% based on t-statistics.

Table III shows the forecast coverage rate at all horizons in both estimating schemes. In the recursive scheme, the performance of individual models (models 1–16) is mixing, i.e. most of the models with asymmetric GARCH each have a correct coverage (the exceptions are models 11 and 12), whereas the five models with symmetric GARCH (models 1, 3, 4, 5, 8) each have an incorrect coverage. This result seems to suggest that the inclusion of asymmetric GARCH in the models improves density forecast accuracy. The combination of densities proposed by Hall and Mitchell (2007) works well when one combines densities only from asymmetric GARCH models (models 18, 21 and 24), although its performance at $h = 2$ is unsatisfactory. We now turn to the quantile methods proposed in this paper, recalling that they are combining point forecasts rather than density forecasts. Model 26 is the location model in which only the intercept changes across quantiles, representing a special case of model 27. Table III shows that model 26 has poor coverage, whereas model 27 has a correct one at all h, suggesting that allowing point forecasts to affect both location and scale helps improve density forecasts. [16]

The results for the rolling window scheme are somewhat similar, showing a mixed performance of individual models but superior performance of symmetric GARCH models over the asymmetric ones.[17] The performance of the quantile method (model 27) is not worse than the existing methods, although it presents a conservative coverage at long horizons such as $h = 12$. The quantile combination method proposed by Granger (model 28) also works very well, presenting a performance as good as model 27. However, since model 27 is based on a more restrictive set of information (point forecasts) its good coverage rate suggests that the approach proposed in this paper can be seen as a good

[16] The null hypothesis of location model was rejected against a location-scale model at 1% by the Koenker and Machado (1999) test.

[17] This result may suggest that the small sample used in the rolling window is affecting the performance of asymmetric GARCH models.

DOI: 10.1002/jae

Table III. Real-time forecast coverage rates (frequencies of actual outcomes falling inside 90% interval band)

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Recursive estimation																												
$h=1$	0.92	0.88	0.90	0.85	0.88	0.85	0.83	0.92	0.89	0.86	0.89	0.86	0.90	0.86	0.87	0.90	0.91	0.92	0.92	0.91	0.92	0.92	0.91	0.92	0.92	0.99	0.89	0.89
	(0.51)	(0.54)	(0.97)	(0.17)	(0.54)	(0.17)	(0.07)	(0.51)	(0.74)	(0.26)	(0.74)	(0.26)	(0.97)	(0.26)	(0.38)	(0.97)	(0.77)	(0.51)	(0.51)	(0.77)	(0.51)	(0.51)	(0.77)	(0.51)	(0.51)	(0)	(0.74)	(0.74)
$h=2$	0.94	0.91	0.94	0.93	0.94	0.88	0.87	0.89	0.94	0.90	0.94	0.94	0.93	0.91	0.91	0.87	0.95	0.95	0.96	0.95	0.95	0.96	0.95	0.95	0.95	1.00	0.93	0.93
	(0.04)	(0.77)	(0.14)	(0.3)	(0.14)	(0.56)	(0.4)	(0.76)	(0.11)	(0.98)	(0.03)	(0.06)	(0.34)	(0.77)	(0.78)	(0.37)	(0.01)	(0.01)	(0)	(0.01)	(0.01)	(0)	(0.01)	(0.01)	(0.01)	(0)	(0.27)	(0.54)
$h=3$	0.94	0.91	0.96	0.96	0.95	0.93	0.92	0.96	0.93	0.90	0.97	0.94	0.92	0.93	0.92	0.93	0.95	0.93	0.95	0.95	0.93	0.94	0.95	0.93	0.94	0.96	0.87	0.96
	(0.16)	(0.76)	(0)	(0)	(0.01)	(0.34)	(0.54)	(0)	(0.35)	(0.97)	(0)	(0.04)	(0.55)	(0.26)	(0.47)	(0.31)	(0.01)	(0.26)	(0.01)	(0.01)	(0.26)	(0.04)	(0.01)	(0.26)	(0.12)	(0)	(0.42)	(0.05)
$h=4$	0.92	0.93	0.96	0.91	0.89	0.87	0.88	0.93	0.90	0.90	0.97	0.89	0.89	0.89	0.90	0.90	0.93	0.92	0.93	0.93	0.92	0.93	0.93	0.90	0.93	0.92	0.82	0.94
	(0.59)	(0.38)	(0)	(0.8)	(0.78)	(0.46)	(0.59)	(0.41)	(0.98)	(0.98)	(0)	(0.78)	(0.78)	(0.78)	(0.98)	(0.98)	(0.41)	(0.61)	(0.38)	(0.41)	(0.61)	(0.38)	(0.41)	(0.98)	(0.38)	(0.79)	(0.06)	(0.98)
$h=8$	0.96	0.94	0.97	0.94	0.86	0.90	0.82	0.94	0.94	0.92	0.94	0.90	0.84	0.84	0.79	0.91	0.94	0.90	0.93	0.94	0.91	0.92	0.94	0.90	0.92	0.94	0.93	0.87
	(0.02)	(0.28)	(0)	(0.23)	(0.54)	(0.98)	(0.23)	(0.31)	(0.35)	(0.62)	(0.07)	(0.98)	(0.32)	(0.25)	(0.06)	(0.85)	(0.29)	(0.99)	(0.51)	(0.29)	(0.85)	(0.71)	(0.29)	(0.99)	(0.71)	(0.38)	(0.56)	(0.35)
$h=12$	0.92	0.84	0.99	0.83	0.83	0.83	0.83	0.89	0.91	0.82	0.98	0.82	0.82	0.82	0.84	0.85	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.90	0.87	0.90
	(0.69)	(0.5)	(0)	(0.45)	(0.47)	(0.43)	(0.43)	(0.87)	(0.85)	(0.37)	(0)	(0.29)	(0.34)	(0.26)	(0.48)	(0.49)	(0.87)	(0.87)	(0.87)	(0.87)	(0.87)	(0.87)	(0.87)	(0.87)	(0.87)	(0.98)	(0.23)	(0.84)
Rolling window																												
$h=1$	0.86	0.74	0.81	0.76	0.76	0.77	0.74	0.77	0.82	0.75	0.81	0.77	0.80	0.77	0.77	0.72	0.83	0.81	0.84	0.83	0.81	0.83	0.83	0.81	0.83	1.00	0.89	0.87
	(0.26)	(0)	(0.02)	(0)	(0)	(0)	(0)	(0)	(0.03)	(0)	(0.02)	(0)	(0.01)	(0)	(0)	(0)	(0.07)	(0.02)	(0.11)	(0.07)	(0.02)	(0.07)	(0.07)	(0.02)	(0.07)	(0)	(0.74)	(0.26)
$h=2$	0.89	0.83	0.88	0.79	0.76	0.73	0.73	0.83	0.86	0.80	0.84	0.75	0.74	0.78	0.74	0.81	0.84	0.79	0.82	0.84	0.81	0.81	0.83	0.80	0.80	1.00	0.88	0.92
	(0.76)	(0.07)	(0.6)	(0.01)	(0)	(0)	(0)	(0.09)	(0.27)	(0.01)	(0.13)	(0)	(0)	(0.01)	(0)	(0.03)	(0.14)	(0.01)	(0.05)	(0.14)	(0.03)	(0.03)	(0.06)	(0.01)	(0.01)	(0)	(0.56)	(0.75)
$h=3$	0.90	0.87	0.91	0.84	0.79	0.74	0.76	0.81	0.86	0.84	0.94	0.82	0.81	0.78	0.75	0.81	0.91	0.91	0.89	0.89	0.89	0.87	0.88	0.89	0.87	0.96	0.91	0.94
	(0.98)	(0.4)	(0.81)	(0.18)	(0.02)	(0.01)	(0.01)	(0.06)	(0.38)	(0.14)	(0.14)	(0.05)	(0.04)	(0.03)	(0.01)	(0.06)	(0.78)	(0.79)	(0.76)	(0.75)	(0.76)	(0.41)	(0.57)	(0.76)	(0.41)	(0.06)	(0.68)	(0.16)
$h=4$	0.89	0.84	0.92	0.80	0.75	0.71	0.72	0.84	0.84	0.85	0.92	0.79	0.75	0.71	0.70	0.81	0.83	0.84	0.84	0.83	0.84	0.84	0.83	0.83	0.84	0.94	0.83	0.91
	(0.79)	(0.23)	(0.59)	(0.04)	(0.01)	(0)	(0)	(0.26)	(0.31)	(0.33)	(0.56)	(0.03)	(0)	(0)	(0)	(0.09)	(0.18)	(0.19)	(0.17)	(0.12)	(0.19)	(0.17)	(0.12)	(0.13)	(0.17)	(0.63)	(0.05)	(0.98)
$h=8$	0.87	0.77	0.92	0.82	0.74	0.74	0.71	0.86	0.85	0.77	0.87	0.78	0.74	0.70	0.69	0.85	0.86	0.83	0.85	0.85	0.83	0.84	0.85	0.82	0.84	0.89	0.84	0.89
	(0.62)	(0.09)	(0.65)	(0.19)	(0.08)	(0.02)	(0)	(0.52)	(0.51)	(0.08)	(0.58)	(0.06)	(0.07)	(0)	(0)	(0.47)	(0.49)	(0.26)	(0.44)	(0.39)	(0.26)	(0.35)	(0.39)	(0.23)	(0.35)	(0.59)	(0.14)	(0.43)
$h=12$	0.84	0.72	0.95	0.75	0.66	0.72	0.72	0.83	0.83	0.66	0.94	0.71	0.65	0.69	0.68	0.80	0.83	0.80	0.81	0.83	0.81	0.82	0.83	0.81	0.81	0.81	0.79	0.83
	(0.45)	(0.07)	(0.16)	(0.11)	(0.02)	(0.02)	(0.02)	(0.34)	(0.34)	(0.01)	(0.22)	(0.04)	(0.02)	(0.01)	(0.01)	(0.24)	(0.4)	(0.25)	(0.28)	(0.4)	(0.28)	(0.32)	(0.4)	(0.28)	(0.28)	(0.15)	(0.05)	(0.09)

Note: The table includes in parentheses p-values for the null of correct coverage (empirical = nominal rate of 90%), based on t-statistics using standard errors computed with the Newey–West estimator, with a bandwidth of 0 at the one-quarter horizon and 1.5× horizon for other horizons.

complement to the existing methods when individual models (and therefore densities and quantiles) are unknown and only point forecasts (limited information) are available.

In what follows we report results obtained by using log scores, which are a broader measure of density calibration.

Table IV reports the values of the log predictive density scores (LPDS), noting that models with higher scores present the best performance. In general, the method proposed by Hall and Mitchell (2007) improves forecast accuracy over some but not all individual models. As suggested by Kascha and Ravazzolo (2010), this result can be interpreted as an insurance against bad models provided by the linear combination (equation (6)). Such a performance, however, is still affected by the fact that all individual models assume a parametric form for the distribution function, which can be different from the true one. Indeed, we also observe that no other model has a LPDS that is higher than that of model 27 at both estimating schemes and at all forecasting horizons. The quantile combination method (model 28) outperforms the individual models and the combined densities, but does not seem to outperform the proposed approach (model 27). To help provide a rough gauge of the significance of score differences, we rely on the methodology developed in Amisano and Giacomini (2007), and report p-values for differences between the LPDS of model 27 and the other models, under the null of equal LPDS. Because the theoretical basis for the test provided by Amisano and Giacomini requires forecasts estimates with rolling samples of data, we only apply the test to the models estimated with the rolling scheme.

Table V shows the results of the Amisano–Giacomini test. A p-value lower than 0.05 indicates that the null hypothesis of equal LPDS between model 27 and model i $(i \neq 27)$ is rejected at a 5% level. Based on the p-values reported in Table V, we can conclude that almost no other model has an LPDS that is statistically equal to the LPDS of model 27, reinforcing our previews results about the good performance of the proposed approach. In sum, our empirical exercise indicates that our combination method is rarely outperformed by either any of the individual models or by combinations of densities and quantiles obtained using equations (6) and (9), respectively. In this sense, we believe that this research fills an important gap in this literature by providing a simple but efficient tool to construct optimal density forecasts without requiring complete information on the individual econometric models. To the best of our knowledge, no other paper has fully explored this possibility.

Finally, it is important to show how the proposed method assigns different weights to the component forecasts. We recall that density forecasts benefit from three important aspects that are taken into account by our approach. First, our model allows for time-varying volatility, which is important especially for forecasting density of inflation since it is well known that the variance of inflation was very low during the Great Moderation period but increased recently due to the sub-prime economic crisis. Thus using models that do not account for time-varying volatilities will result in inaccurate density forecasts. Second, most of the existing literature on density forecasting take a stand on the functional form of the forecasting error distribution. If such a specification is wrong, then density forecasts will also be inaccurate. Our approach, on the other hand, relies on QR techniques which do not need to assume the functional form of the error distribution. Third, we minimize the risk of misspecification in the quantile function through the combination of point forecasts. This same technique was used by Granger and Ramanathan (1984) to minimize the misspecification risk in the mean function. In this paper, using point forecasts to approximate the true quantile function has a natural interpretation as a generalization of the result of Granger and Ramanathan (1984) from point to density forecasting.

Figures 3 and 4 show the weights obtained from our quantile method, which are nothing but an intercept and slope coefficients from a quantile regression of y_{t+h} on a constant and eight point forecasts. To save space we set $h=3$ and $\tau=0.75, 0.5$ and 0.25. The rolling window scheme is used. Figure 3 shows optimal QR weights for $\tau=0.75$, which is a relevant measure of upside risk of inflation (i.e. inflation values above median). One can see that from 1985 to 1988 (period of volatile inflation) the weights are quite different, suggesting that forecast combination plays an important role in forecasting the upside risk of inflation. During the Great

DOI: 10.1002/jae

Table IV. Log predictive density score (LPDS)

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Recursive estimation																												
$h=1$	−3.32	−3.31	−3.32	−3.28	−3.29	−3.27	−3.26	−3.27	−3.32	−3.31	−3.31	−3.28	−3.29	−3.26	−3.26	−3.27	−3.29	−3.29	−3.28	−3.29	−3.29	−3.28	−3.29	−3.29	−3.28	−3.34	−3.03	−3.15
$h=2$	−3.36	−3.33	−3.36	−3.35	−3.34	−3.32	−3.32	−3.31	−3.36	−3.33	−3.35	−3.34	−3.33	−3.31	−3.31	−3.31	−3.34	−3.33	−3.33	−3.34	−3.33	−3.33	−3.34	−3.33	−3.33	−3.30	−3.04	−3.25
$h=3$	−3.39	−3.35	−3.41	−3.37	−3.36	−3.36	−3.36	−3.38	−3.38	−3.35	−3.41	−3.36	−3.35	−3.35	−3.34	−3.37	−3.37	−3.37	−3.36	−3.37	−3.36	−3.36	−3.37	−3.36	−3.36	−3.25	−2.98	−3.24
$h=4$	−3.41	−3.41	−3.46	−3.39	−3.38	−3.39	−3.38	−3.42	−3.41	−3.40	−3.46	−3.38	−3.38	−3.39	−3.38	−3.42	−3.41	−3.41	−3.40	−3.40	−3.40	−3.40	−3.40	−3.40	−3.39	−3.43	−2.98	−3.05
$h=8$	−3.48	−3.49	−3.58	−3.47	−3.44	−3.47	−3.44	−3.48	−3.47	−3.48	−3.57	−3.47	−3.43	−3.47	−3.43	−3.47	−3.48	−3.48	−3.48	−3.48	−3.47	−3.48	−3.48	−3.47	−3.47	−3.53	−3.07	−3.29
$h=12$	−3.51	−3.52	−3.70	−3.51	−3.51	−3.53	−3.53	−3.50	−3.51	−3.51	−3.70	−3.51	−3.51	−3.52	−3.52	−3.50	−3.53	−3.53	−3.53	−3.53	−3.54	−3.53	−3.53	−3.53	−3.53	−3.52	−3.08	−3.39
Rolling window																												
$h=1$	−3.26	−3.22	−3.27	−3.19	−3.21	−3.15	−3.13	−3.20	−3.25	−3.21	−3.26	−3.17	−3.20	−3.14	−3.14	−3.19	−3.21	−3.20	−3.19	−3.21	−3.20	−3.20	−3.21	−3.19	−3.19	−3.31	−3.04	−3.14
$h=2$	−3.31	−3.27	−3.34	−3.24	−3.22	−3.23	−3.21	−3.27	−3.31	−3.26	−3.33	−3.23	−3.21	−3.20	−3.19	−3.26	−3.26	−3.25	−3.25	−3.26	−3.26	−3.25	−3.26	−3.25	−3.24	−3.26	−2.96	−3.22
$h=3$	−3.35	−3.33	−3.38	−3.28	−3.21	−3.24	−3.24	−3.30	−3.36	−3.32	−3.38	−3.27	−3.20	−3.25	−3.23	−3.30	−3.29	−3.28	−3.28	−3.28	−3.28	−3.27	−3.28	−3.28	−3.27	−3.32	−3.09	−3.20
$h=4$	−3.38	−3.33	−3.44	−3.28	−3.24	−3.28	−3.22	−3.36	−3.37	−3.33	−3.43	−3.28	−3.24	−3.27	−3.22	−3.36	−3.33	−3.32	−3.32	−3.32	−3.32	−3.31	−3.32	−3.31	−3.31	−3.48	−3.07	−3.05
$h=8$	−3.42	−3.43	−3.56	−3.41	−3.35	−3.41	−3.40	−3.43	−3.43	−3.42	−3.55	−3.41	−3.34	−3.40	−3.40	−3.43	−3.43	−3.43	−3.43	−3.43	−3.43	−3.42	−3.42	−3.42	−3.42	−3.51	−3.07	−3.28
$h=12$	−3.45	−3.38	−3.68	−3.40	−3.37	−3.38	−3.36	−3.44	−3.44	−3.37	−3.68	−3.39	−3.37	−3.36	−3.34	−3.43	−3.46	−3.45	−3.46	−3.46	−3.46	−3.46	−3.45	−3.44	−3.45	−3.57	−3.07	−3.30

Note: The table entries are average values of log predictive density scores (see Adolfson *et al*., 2005), under which a higher score implies a better model.

Table V. Amisano-Giacomini (2007) test applied to average LPDS

Rolling window

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
$h=1$	0.22	0.18	0.23	0.15	0.17	0.11	0.10	0.16	0.22	0.17	0.22	0.14	0.16	0.10	0.10	0.15	0.17	0.16	0.16	0.17	0.16	0.16	0.17	0.16	0.15	0.28	–	0.10
	(0)	(0)	(0)	(0.01)	(0)	(0.04)	(0.07)	(0)	(0)	(0)	(0)	(0.01)	(0)	(0.05)	(0.06)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)		(0.01)
$h=2$	0.35	0.31	0.37	0.28	0.26	0.27	0.25	0.31	0.34	0.30	0.37	0.27	0.25	0.24	0.23	0.29	0.30	0.29	0.28	0.30	0.29	0.28	0.30	0.29	0.28	0.30	–	0.26
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)		(0)
$h=3$	0.26	0.23	0.29	0.19	0.12	0.15	0.14	0.21	0.26	0.23	0.28	0.18	0.11	0.15	0.14	0.20	0.20	0.19	0.19	0.19	0.19	0.18	0.19	0.18	0.18	0.23	–	0.11
	(0)	(0)	(0)	(0.01)	(0.15)	(0.04)	(0.05)	(0)	(0)	(0)	(0)	(0.01)	(0.16)	(0.04)	(0.06)	(0)	(0)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0)		(0.21)
$h=4$	0.30	0.26	0.36	0.20	0.16	0.21	0.15	0.29	0.30	0.26	0.36	0.20	0.16	0.20	0.15	0.29	0.25	0.25	0.24	0.25	0.25	0.24	0.25	0.24	0.23	0.41	–	−0.02
	(0)	(0)	(0)	(0)	(0.01)	(0)	(0.03)	(0)	(0)	(0)	(0)	(0)	(0.01)	(0)	(0.03)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)		(0.8)
$h=8$	0.35	0.36	0.48	0.34	0.28	0.34	0.33	0.36	0.36	0.35	0.48	0.34	0.27	0.33	0.33	0.36	0.36	0.36	0.35	0.36	0.36	0.35	0.35	0.35	0.35	0.44	–	0.21
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)		(0.03)
$h=12$	0.39	0.31	0.61	0.33	0.30	0.31	0.29	0.37	0.38	0.30	0.61	0.32	0.30	0.29	0.27	0.37	0.39	0.38	0.39	0.40	0.39	0.39	0.38	0.37	0.38	0.50	–	0.24
	(0.01)	(0.06)	(0)	(0.04)	(0.07)	(0.03)	(0.04)	(0.01)	(0.01)	(0.07)	(0)	(0.05)	(0.07)	(0.04)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0)		(0.13)

Note: Null hypothesis of zero average difference in LPDS between model 27 (benchmark) and model $i \neq 27$. Similar to Clark (2011), the p-values (in parentheses) are computed by regressions of differences in log scores (time series) on a constant, using the Newey–West estimator of the variance of the regression constant (with a bandwidth of 0 at the one-quarter horizon and 1.5× horizon for other horizons).

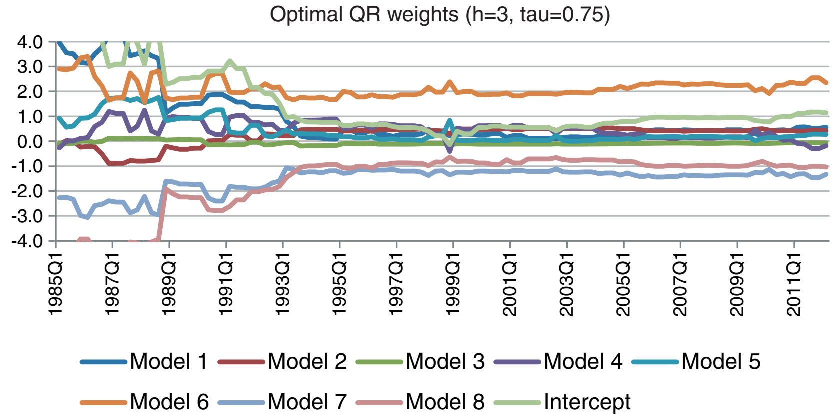

Figure 3. Point forecast combination $\tau = 0.75$

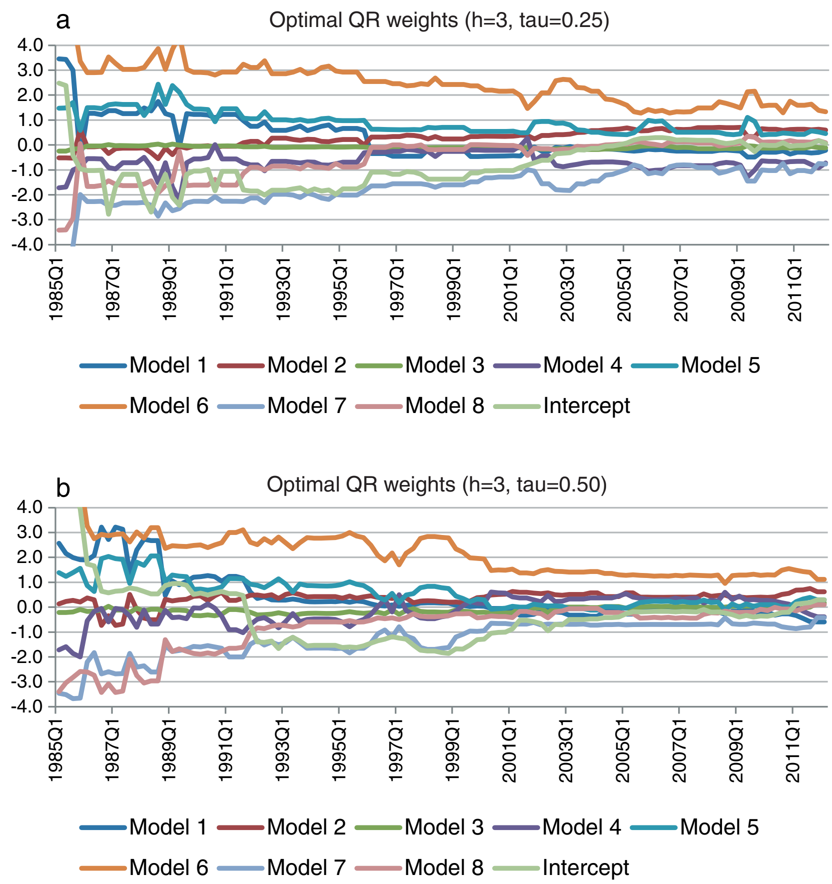

Figure 4. Point forecast combination $\tau = 0.25$ and 0.50

DOI: 10.1002/jae

Moderation (1988–2008), inflation was low, less volatile and therefore easier to forecast by Phillips curve models with survey expectations. As a result, weights on some models approach zero and forecast combination plays a less important role in predicting the upside risk. Finally, at the time the mortgage crises erupted (2008–2012), the volatility of g_t increased and the weights on some models started to be different from zero again, making forecast combination important in predicting the upside risk of inflation. One explanation for this result is that the coefficients of the QR (optimal weights) depend on $F_{\eta,h}^{-1}(\tau_i)$, $\tau_i = 0.75$, which is left unspecified, and therefore it is able to adapt quickly to changes caused by economic events.

Figure 4 shows the QR weights for $\tau = 0.5$ and $\tau = 0.25$, which are less important in predicting the upside risk of inflation but are important in predicting its central tendency and downside risk (i.e. values of inflation below median). Recall that these weights depend on $F_{\eta,h}^{-1}(\tau_i)$, $\tau_i = 0.5$ and 0.25, respectively. One can observe that the weights are highly disperse during the period of high inflation but more stable during the rest of the sample, i.e. $F_{\eta,h}^{-1}(\tau_i)$, $\tau_i = 0.5$ and 0.25 did not change much from 1988 to 2012. Thus the effect of the mortgage crisis seems not to be captured at low quantiles since the QR weights are very similar to the values observed during the Great Moderation. This may be explained by the fact that the current mortgage crisis had a deflationary impact on the US economy, making it easier to forecast the central tendency and the downside risk of inflation, i.e. less dispersion in the left tail of the inflation distribution. Finally, based on Figures 3 and 4, models with survey expectations (i.e. models 6, 7 and 8) seem to dominate other models throughout the time examined, indicating that survey expectations play an important role in the prediction of density forecasts.

6. CONCLUSION

Granger and Ramanathan (1984) advocated that the combination of point forecasts should be used when there is uncertainty about the true specification of the conditional mean function. In this paper, we show that the same idea can be employed to mitigate uncertainty about the true specification of the conditional quantile function. This result can be applied to construct density forecasts when the decision maker has limited information, i.e. when he or she observes the point forecasts computed by economic institutions but does not observe the econometric models used by them. Under this situation, the combination devices proposed by Hall and Mitchell (2007) and Granger (1969, 1989), which are based on full knowledge of the unknown econometric models, are no longer feasible.

The methodology developed in this paper provides a simple and efficient way to estimate the uncertainty behind economic forecasting and therefore can be useful in identifying the correct economic policy under different circumstances. Perhaps most importantly, our approach is applicable under a wide variety of structures, since it does not require full knowledge of the unknown econometric models, including the specification of the forecast error distribution function. Given this approach, we were able to make h -step-ahead forecasts of any quantile of y_{t+h} and therefore forecast the entire density.

We provide empirical evidence of our theoretical findings by forecasting the density of future inflation in the USA. There has been intense research on forecasting the behavior of inflation, but most of the papers focus on modeling the conditional mean or the most likely outcome. If the decision maker is interested in evaluating the upside or downside risks of inflation, then a forecast of the density, $f_{t+h,t}$, is necessary. In this paper, we use our proposed approach to estimate $f_{t+h,t}$ and compare it to density forecasts obtained from existing methods. The evidence presented in this paper shows that the proposed optimal combination method materially improves the real-time accuracy of density forecasts. The empirical evidence includes interval forecasts (coverage rates) and log predictive density scores.

Although our empirical results are favorable, we are not claiming that our method will always outperform the combination method suggested by Hall and Mitchell (2007) and Granger (1969, 1989). Our main contribution is to show that accurate density forecasts can be obtained even when we do not have full knowledge about the specification of individual models. Under this limited information setting, our approach can be interpreted as a complement to the existing ones, without ruling out the possibility that other individual models could be included in $\mathcal{F}_t$.

ACKNOWLEDGEMENTS

The authors thank the Editor M. Hashem Pesaran, Co-Editor Francis X. Diebold, two anonymous referees, Roger Koenker, Todd E. Clark, Zhijie Xiao, Williams Nielson and seminar participants in the University of Tennessee, University of Illinois at Urbana-Champaign, EESP-FGV, 32nd International Symposium on Forecasting, for their helpful comments and suggestions. Lima thanks CNPq for financial support. The opinions expressed in this article are those of the authors and do not necessarily reflect those of the Banco Central do Brasil. Any remaining errors are ours.

REFERENCES

Adolfson M, Linde J, Villani M. 2005. Forecasting performance of an open economy dynamic stochastic general equilibrium model. Sveriges Riksbank Working Paper 190.

Amisano G, Giacomini R. 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics* **25**: 177–190.

Andrade P, Ghysels E, Idier J. 2012. Tails of inflation forecasts and tales of monetary policy. Working Paper 407, Banque de France.

Atkeson A, Ohanian LE. 2001. Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review* **25**: 2–11.

Britton E, Fisher P, Whitley J. 1998. The Inflation Report projections: understanding the fan chart. *Bank of England Quarterly Bulletin* **38**(1): 30.

Carriero A, Clark T, Marcellino M. 2012a. Real time nowcasting with a Bayesian mixed frequency model with stochastic volatility. Cleveland FED Working Paper 12-27.

Carriero A, Clark T, Marcellino M. 2012b. Common drifting volatility in large Bayesian VARs. CEPR Working Paper 8894.

Christoffersen PF, Diebold FX. 1997. Optimal prediction under asymmetric loss. *Econometric Theory* **13**: 808–817.

Clark TE. 2011. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics* **29**: 327–341.

Clemen RT. 1989. Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* **5**: 559–583.

Elliott G, Timmermann A. 2004. Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics* **122**: 47–79.

Elliott G, Timmermann A. 2008. Economic Forecasting. *Journal of Economic Literature* **46**(1): 3–56.

Faust J, Wright JH. 2012. Forecasting inflation. Johns Hopkins University working paper.

Friedman M. 1977. Nobel lecture: inflation and unemployment. *Journal of Political Economy* **85**(3): 451–472.

Gaglianone WP, Lima, LR. 2012. Constructing density forecasts from quantile regression. *Journal of Money, Credit, and Banking* **44**(8): 1589–1607.

Giordani P, Villani M. 2010. Forecasting macroeconomic time series with locally adaptive signal extraction. *International Journal of Forecasting* **26**(2): 312–325.

Granger CWJ. 1969. Prediction with a generalized cost of error function. *Operational Research Quarterly* **20**(2): 199–207.

Granger CWJ. 1989. Combining forecasts: twenty years later. *Journal of Forecasting* **8**: 167–173.

Granger CWJ, Newbold P. 1986. *Forecasting time series* (2nd edn). Academic Press: New York.

Granger CWJ, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* **3**: 197–204.

Hall SG, Mitchell J. 2007. Combining density forecasts. *International Journal of Forecasting* **23**: 1–13.

Issler JV, Lima LR. 2009. A panel data approach to economic forecasting: the bias-corrected average forecast. *Journal of Econometrics* **152**: 153–164.

Jore AS, Mitchell J, Vahey SP. 2010. Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics* **25**(4): 621–634.

Judson R, Orphanides A. 1999. Inflation, volatility and growth. *International Finance* **2**: 117–138.
Kascha C, Ravazzolo F. 2010. Combining inflation density forecasts. *Journal of Forecasting* **29**: 231–250.
Koenker R. 2005. *Quantile Regression*. Cambridge University Press: Cambridge, UK.
Koenker R, Basset GW. 1978. Regression quantiles. *Econometrica* **46**(1): 33–50.
Koenker R, Machado JAF. 1999. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* **94**: 1296–1310.
Marcellino M. 2004. Forecast pooling for short time series of macroeconomic variables. *Oxford Bulletin of Economic and Statistics* **66**: 91–112.
Marcellino M, Stock J, Watson M. 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* **135**: 499–526.
Mitchell J, Wallis KF. 2010. Evaluating density forecasts: forecast combinations, model mixtures, calibration, and sharpness. *Journal of Applied Econometrics* **26**(6): 1023–1040.
Patton AJ, Timmermann A. 2007. Testing forecast optimality under unknown loss. *Journal of the American Statistical Association* **102**(480): 1172–1184.
Stock JH, Watson MW. 2001. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In *Festschrift in Honour of Clive Granger*, Engle RF, White H (eds). Cambridge University Press: Cambridge, UK; 1–44.
Stock JH, Watson MW. 2004. Combining forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23**: 405–430.
Stock JH, Watson MW. 2008. Phillips curve inflation forecasts. NBER Working Paper 14322.
Stone M. 1961. The opinion pool. *Annals of Mathematical Statistics* **32**: 1339–1342.
Timmermann A. 2006. Forecast combinations. In *Handbook of Economic Forecasting*, Vol. **1**, Elliott G, Granger CWJ, Timmermann A (eds). Elsevier: Amsterdam; 135–154.

APPENDIX

Proof of Proposition 1. The proof is similar to that shown by Granger (1969), Christoffersen and Diebold (1997) and Patton and Timmermann (2007) in the first part of their Proposition 2. Thus, by homogeneity of the loss function and DGP (1), we have that

$$
\begin{aligned}
\hat{y}^i_{t+h,t} &= \arg\min_{\hat{y}} \int L^i(y-\hat{y})dF_{t+h,t}(y) \\
&= \arg\min_{\hat{y}} \int \left[g\left(\frac{1}{\hat{\mathbf{y}}'_{t+h,t}\gamma} \right) \right]^{-1} L^i\left(\frac{1}{\hat{\mathbf{y}}'_{t+h,t}\gamma}(y-\hat{y}) \right) dF_{t+h,t}(y) \\
&= \arg\min_{\hat{y}} \int \left[g\left(\frac{1}{\left(\gamma_0+\gamma_1\hat{y}^1_{t+h,t}+\ldots+\gamma_{k-1}\hat{y}^{k-1}_{t+h,t}\right)} \right) \right]^{-1} L^i\left(\frac{1}{\left(\gamma_0+\gamma_1\hat{y}^1_{t+h,t}+\ldots+\gamma_{k-1}\hat{y}^{k-1}_{t+h,t}\right)}(y-\hat{y}) \right) dF_{t+h,t}(y) \\
&= \arg\min_{\hat{y}} \int L^i\left(\frac{1}{\left(\gamma_0+\gamma_1\hat{y}^1_{t+h,t}+\ldots+\gamma_{k-1}\hat{y}^{k-1}_{t+h,t}\right)}(y-\hat{y}) \right) dF_{t+h,t}(y) \\
&= \arg\min_{\hat{y}} \int L^i\left(\begin{array}{c} \dfrac{1}{\left(\gamma_0+\gamma_1\hat{y}^1_{t+h,t}+\ldots+\gamma_{k-1}\hat{y}^{k-1}_{t+h,t}\right)}\cdot \\ \cdot\left(\begin{array}{c} \omega_0+\omega_1\hat{y}^1_{t+h,t}+\ldots+\omega_{k-1}\hat{y}^{k-1}_{t+h,t}+\gamma_0\eta_{t+h}+ \\ +\gamma_1\hat{y}^1_{t+h,t}\eta_{t+h}+\ldots+\gamma_{k-1}\hat{y}^{k-1}_{t+h,t}\eta_{t+h}-\hat{y} \end{array} \right) \end{array} \right) dF_{t+h,t}(y)
\end{aligned}
$$

Let us represent a forecast by

$$\omega_0 + \omega_1 \hat{y}^1_{t+h,t} + \ldots + \omega_{k-1} \hat{y}^{k-1}_{t+h,t} + \left(\gamma_0 + \gamma_1 \hat{y}^1_{t+h,t} + \ldots + \gamma_{k-1} \hat{y}^{k-1}_{t+h,t}\right) \hat{\gamma}_{t+h,t}$$

This way, it follows that

$$\hat{y}^i_{t+h,t} = \omega_0 + \omega_1 \hat{y}^1_{t+h,t} + \ldots + \omega_{k-1} \hat{y}^{k-1}_{t+h,t} + \left(\gamma_0 + \gamma_1 \hat{y}^1_{t+h,t} + \ldots + \gamma_{k-1} \hat{y}^{k-1}_{t+h,t}\right)$$

$$\cdot \arg\min_{\hat{\gamma}} \int L^i\left(\frac{1}{\left(\gamma_0 + \gamma_1 \hat{y}^1_{t+h,t} + \ldots + \gamma_{k-1} \hat{y}^{k-1}_{t+h,t}\right)}\left(\omega_0 + \omega_1 \hat{y}^1_{t+h,t} + \ldots\right.\right.$$

$$+\omega_{k-1} \hat{y}^{k-1}_{t+h,t} + \left(\gamma_0 + \gamma_1 \hat{y}^1_{t+h,t} + \ldots + \gamma_{k-1} \hat{y}^{k-1}_{t+h,t}\right)\eta_{t+h} - \omega_0 - \omega_1 \hat{y}^1_{t+h,t} - \ldots$$

$$\left.\left.-\omega_{k-1} \hat{y}^{k-1}_{t+h,t} - \left(\gamma_0 + \gamma_1 \hat{y}^1_{t+h,t} + \ldots + \gamma_{k-1} \hat{y}^{k-1}_{t+h,t}\right)\hat{\gamma}\right)\right) \mathrm{d}F_{\eta,h}(\eta)$$

$$= \omega_0 + \omega_1 \hat{y}^1_{t+h,t} + \ldots + \omega_{k-1} \hat{y}^{k-1}_{t+h,t} +$$

$$+\left(\gamma_0 + \gamma_1 \hat{y}^1_{t+h,t} + \ldots + \gamma_{k-1} \hat{y}^{k-1}_{t+h,t}\right) \cdot \arg\min_{\hat{\gamma}} \int L^i\left(\eta_{t+h} - \hat{\gamma}\right) \mathrm{d}F_{\eta,h}(\eta)$$

$$= \omega_0 + \gamma_0 \gamma^i_h + \omega_1 \hat{y}^1_{t+h,t} + \gamma_1 \hat{y}^1_{t+h,t} \gamma^i_h + \ldots + \omega_{k-1} \hat{y}^{k-1}_{t+h,t} + \gamma_{k-1} \hat{y}^{k-1}_{t+h,t} \gamma^i_h$$

$$= \omega_0(\tau_i) + \omega_1(\tau_i) \hat{y}^1_{t+h,t} + \ldots + \omega_{k-1}(\tau_i) \hat{y}^{k-1}_{t+h,t}$$

where $\omega_0(\tau_i) = \left(\omega_0 + \gamma_0 \gamma^i_h\right), \quad \omega_j(\tau_i) = (\omega_j + \gamma_j \gamma^i_h), \quad j = 1, \ldots, k-1$

□

Proof of Corollaries 1 and 2. If we assume that there are no scale effects then $\gamma_1 = \ldots = \gamma_{k-1} = 0$ and therefore the optimal forecast will be $\hat{y}^i_{t+h,t} = \omega_0(\tau_i) + \omega_1 \hat{y}^1_{t+h,t} + \ldots + \omega_{k-1} \hat{y}^{k-1}_{t+h,t}$, where $\omega_0(\tau_i) = \left(\omega_0 + \gamma_0 \gamma^i_h\right)$. This proves Corollary 2. To prove Corollary 1 we remember that the expected value of the MSE loss is $\hat{y}^i_{t+h,t} = \arg\min_{\hat{y}} E(y - \hat{y})^2$. Now, due to certainty equivalence $E(y - \hat{y})^2$ is minimized at $\hat{y} = E(y|\hat{y}) = \omega_0 + \omega_1 \hat{y}^1_{t+h,t} + \ldots + \omega_{k-1} \hat{y}^{k-1}_{t+h,t}$. This proves Corollary 1.

□

DOI: 10.1002/jae