Constructing Optimal Density Forecasts from Point Forecast Combinations.

Wagner Piazza Gaglianone* and Luiz Renato Lima†

February 20, 2012

Abstract

Decision makers often observe point forecasts of the same variable computed, for instance, by commercial banks, IMF, World Bank, but the econometric models used by such institutions are unknown. This paper shows how to use the information available at 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 future inflation in the U.S. and our results indicate that the proposed method materially improves the real-time accuracy of density forecasts vis-à-vis the ones obtained from the (unknown) individual econometric models.

Keywords: forecast combination, quantile regression, density forecast, loss function.

JEL Codes: C13, C14, C51, C53.

*Research Department, Central Bank of Brazil, and Fucape Business School (e-mail: wagner.gaglianone@bcb.gov.br).
†Corresponding author. Department of Economics, University of Tennessee at Knoxville, 527A Stokely Management Center (e-mail: llima@utk.edu). phone: (865) 974-1698. Fax: (865) 974-4601
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 motivation for combining different forecasts arises when the information set of the individual models are 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 widespreaded in such diverse areas as economics, finance, meteorology, political sciences, 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.

Our research is related to the literature of 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 at 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 U.S. 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 U.S. remained extremely low during the 1987-2007 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 the ones constructed from the (unknown) individual econometric models nor by the one obtained using the linear combination suggested by Hall and Mitchell (2007). The advantage of our approach can be explained by the fact that no distributional assumption on the forecast error is required and that the uncertainty about the specification of the conditional quantile function is minimized by using the proposed combination device based on point forecasts. The density evidence includes interval forecasts (coverage rates), tests applied to normal transforms of the probability integral transforms, and log predictive density scores.

This paper is organized as follows: Section 2 presents the forecast combination problem, discusses the econometric model and assumptions, presents our results on optimal forecast combination, and shows how to use the proposed method to construct a density forecast. Section 3 presents our empirical illustration whereas section 4 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 $\hat{y}_{t+h,t} = \left(\hat{y}_{t+h,t}^{1}, \hat{y}_{t+h,t}^{2}, \ldots, \hat{y}_{t+h,t}^{k}\right)'$. Notice that each element of $\hat{y}_{t+h,t}$ is determined ex-ante (at time $t$) and is adapted to an expanding sequence of information sets $\mathcal{F}_t$. Hence, $\hat{y}_{t+h,t}$
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 $y_{t+h}$. We seek an aggregator that reduces the information in $\tilde{\mathbf{y}}_{t+h,t} \in \mathbb{R}^k$ to a summary measure $C(\tilde{\mathbf{y}}_{t+h,t}, \omega^j) \in \mathbb{R}$. Identification of $\omega^j \in \mathbb{R}^k$ will depend on both the general 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 $\hat{f}_{t+h,t}$ of the conditional density (or distribution function) of $y_{t+h}$.

We assume that $y_{t+h}$ and $\tilde{\mathbf{y}}_{t+h,t}$ have stationary joint distribution

$$F(y_{t+h}, \tilde{\mathbf{y}}'_{t+h,t}) = F(y_{t+h} | \tilde{\mathbf{y}}'_{t+h,t}) \cdot F(\tilde{\mathbf{y}}'_{t+h,t})$$

where $F(y_{t+h} | \tilde{\mathbf{y}}'_{t+h,t}) = F_{t+h,t}$ is the conditional distribution of $y_{t+h,t}$ and $F(\tilde{\mathbf{y}}'_{t+h,t})$ is the distribution of $\tilde{\mathbf{y}}'_{t+h,t}$. Since the main interest is to forecast $y_{t+h}$ based on the k-vector of forecasts $\tilde{\mathbf{y}}_{t+h,t}$, we will only model $F_{t+h,t}$. Elliott and Timmermann (2004) derived optimal weights for forecast combination based on the modelling of $F\left(y_{t+h}, \tilde{\mathbf{y}}'_{t+h,t}\right)$, which accounts explicitely for the covariance of the $k$ individual forecasts. 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. However, some of their solutions does not have a closed form and needs to assume knowledge of the data distribution. In this paper we show that most of the results in Elliott and Timmermann (2004) can also be obtained by only modelling $F_{t+h,t}$. 
The conditional model with mean and variance dynamics is defined as

\[
y_{t+h} = \tilde{y}_{t+h,t} + \left( \tilde{y}_{t+h,t} \gamma \right) \eta_{t+h},
\]

\[
\eta_{t+h} | \mathcal{F}_t \sim F_{\eta,h}(0, 1)
\]

where \( F_{\eta,h}(0, 1) \) is some distribution with mean zero and unit variance, which depends on \( h \) but does not depend on \( \mathcal{F}_t \). \( \tilde{y}_{t+h,t} \in \mathcal{F}_t \) is a \( k \times 1 \) vector of forecasts, that can be predicted using information available at time \( t \), and \( \omega \) and \( \gamma \) are \( k \times 1 \) vectors of parameters, which include an intercept. An important thing to notice is that no parametric structure is placed on \( F_{\eta,h} \). In this model, covariates \( \tilde{y}_{t+h,t} \) affect both location and scale of the conditional distribution of \( y_{t+h} \). This class of DGPs is very broad and includes most common volatility processes such as ARCH and stochastic volatility.

Following the literature (i.e. Granger (1969), Granger and Newbold (1986), Christoffersen and Diebold (1997), and Patton and Timmermann (2007)), an optimal forecast combination \( \tilde{y}_{t+h,t} \) 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 \( e_{t+h,t} \), that is,

\[
L_i = L_i(e_{t+h,t}), \quad \text{and} \quad L(ae) = g(a)L(e) \quad \text{for some positive function } g.
\]

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

\[ \hat{y}^{i}_{t+h,t} = \omega_0(\tau_i) + \omega_1(\tau_i)\hat{y}^1_{t+h,t} + \ldots + \omega_k(\tau_i)\hat{y}^k_{t+h,t} \]  

where \( \omega_0(\tau_i) = (\omega_0 + \gamma_0 \gamma^i_h) \); \( \omega_1(\tau_i) = (\omega_1 + \gamma_1 \gamma^i_h) \); \( \omega_k(\tau_i) = (\omega_k + \gamma_k \gamma^i_h) \) and \( \gamma^i_h \) is a constant that depends only on the distribution \( F_{\eta,h}(0,1) \) and the loss function \( L^i \).

**Proof.** See Appendix.

The results in proposition 1 are quite interesting. First, we notice that the optimal weights will depend on the forecast horizon, the loss \( L^i \), and the unknown distribution function \( F_{\eta,h}(0,1) \) through the constant \( \gamma^i_h \). Second, the optimal weights are linear departures from the optimal weights obtained under the MSE loss as we show in the next corollary.

**Corollary 1** Under DGP(1) and the mean-squared-error (MSE) loss function, the optimal forecast combination is

\[ \hat{y}^{i}_{t+h,t} = E[y_{t+h}|F_t] = \omega_0 + \omega_1\hat{y}^1_{t+h,t} + \ldots + \omega_k\hat{y}^k_{t+h,t} \]

where \( E[y_{t+h}|F_t] \) is the conditional mean of \( y_{t+h} \).

**Proof.** See Appendix.

Corollary 1 shows that, under the 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

---

\( ^1 \)Note that \( \omega_j(\tau_i), j = 0, \ldots, k \) are functions of \( \tau_i \) and \( (\omega_j + \gamma_j \gamma^i_h) \) are their respective images.

\( ^2 \)Notice that, by assumption, \( E[y_{t+h}|F_t] = 0 \).
Granger and Ramanathan (1984). Based on basic least-squares estimator algebra, we know that the vector of combination weights depends only on the variance-covariance matrix of $y_{t+h}$ and the predictions $\hat{y}_{t+h,t}^i$, which is essentially the same result reported by Elliott and Timmermann (2004), but obtained using the conditional model (1) rather than their joint distribution approach.

In the absence of scale effects in the conditional model (1), i.e., $\gamma_1 = \gamma_2 = \ldots = \gamma_k = 0$, the optimal forecast combination weights are identical to the MSE weights, and the intercept $\omega_0(\tau_i)$ 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., $\gamma_1 = \gamma_2 = \ldots = \gamma_k = 0$, the optimal forecast combination, under DGP(1), will be

$$\hat{y}_{t+h,t}^i = E[y_{t+h}|F_t] + k^h_i = \omega_0(\tau_i) + \omega_1\hat{y}_{t+h,t}^1 + \ldots + \omega_k\hat{y}_{t+h,t}^k$$

(4)

where $\omega_0(\tau_i) = (\omega_0 + \gamma_0^{\gamma_h})$, and $k^h_i = \gamma_0^{\gamma_h}$

**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 the ones obtained under MSE loss. This same result was also obtained by Elliott and Timmermann (2004) using their 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.

In the presence of location and scale effects, our methodology yields new
results in the sense that the optimal weights will now differ from the ones obtained under MSE loss. The conditional model (1) offers a natural approach to estimating optimal weights without relying on a specific functional form of $F_{\eta,h}(0,1)$ and $L^i$. 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{equation}
F_{t+h,t}(\hat{y}_{t+h,t}^i) = \Pr(y_{t+h} < \hat{y}_{t+h,t}^i | \mathcal{F}_t) = \Pr \left( \begin{array}{l}
y_{t+h} = \hat{y}_{t+h,t}^i \omega + \left( \hat{y}_{t+h,t}^i \gamma \right) \eta_{t+h} < \\
< \hat{y}_{t+h,t}^i \omega + \left( \hat{y}_{t+h,t}^i \gamma \right) \gamma_{t+h} | \mathcal{F}_t
\end{array} \right) = \Pr (\eta_{t+h} < \gamma_{t+h} | \mathcal{F}_t) = F_{\eta,h}(\gamma_{t+h})
\end{equation}

where $F_{\eta,h}(\gamma_{t+h}) = \tau_i$ is a fixed value of $\tau \in (0,1)$. Thus, it follows that, by definition, the optimal forecast combination $\hat{y}_{t+h,t}^i$ must coincide with the conditional quantile function of $y_{t+h}$ at level $\tau_i$, i.e.:

\begin{equation}
\hat{y}_{t+h,t}^i = Q_{y_{t+h}}(\tau_i | \mathcal{F}_t), \text{ for some } \tau_i \in (0,1)
\end{equation}

and

\begin{equation}
Q_{y_{t+h}}(\tau_i | \mathcal{F}_t) = \omega_0(\tau_i) + \omega_1(\tau_i)\hat{y}_{t+h,t}^1 + \ldots + \omega_k(\tau_i)\hat{y}_{t+h,t}^k
\end{equation}

Thus, the optimal weights under general loss $L^i$ and unknown distribution function $F_{\eta,h}(0,1)$ can be obtained through a quantile regression of $y_{t+h}$ on a constant and the vector of forecasts.

**Remark 1** Equation (6) 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 argued by Timmermann (2006), the combined point forecast (which represents a single future path of the target variable) provides an insufficient information for a decision maker who might be interested in the degree of uncertainty surround a point forecast. Indeed, Greenspan (2004, p. 37) discusses his concern about point forecasts in the following terms: "Given our inevitably incomplete knowledge about key structural aspects of an ever-changing economy and the sometimes asymmetric costs or benefits of particular outcomes, a central bank needs to consider not only the most likely future path for the economy, but also the distribution of possible outcomes about the path." For this reason, it would be more correct to forecast the entire density than simply a point forecast. In order to forecast the entire density of \( y_{t+h} \) based on \( \mathcal{F}_t \) we need to be able to identify the entire family of conditional quantiles \( Q_{y_{t+h}}(\tau | \mathcal{F}_t) \), for any \( \tau \in (0,1) \) and therefore all possible paths of the target variable. Figure 1 shows what we are looking for. Indeed, in order to forecast the entire density we would have to be able to identify all possible future paths (the gray area in Figure 1).\(^3\)

\(^3\)The red lines represent the out-of-sample empirical quantiles \( \tau = \{0.25; 0.50; 0.75\} \).
This task is unfeasible if we employ the method proposed by Elliott and Timmermann (2004) because it would lead us to compute the weights for all possible combinations of loss functions and forecast error distributions, a task that may be difficult to implement in practice. Besides, Elliott and Timmermann (2004) show that, for some losses and forecast error distributions, the optimal vector of weights does not even have a closed form solution, which implies that some of the paths in Figure 1 will not be easily identified. Fortunately, we can use the proposed conditional model (1) to identify all possible future paths by noting that they correspond to different conditional quantiles of the target variable. Each conditional quantile $Q_{y_{t+h}}(\tau \mid F_t), \tau \in (0,1)$ can be identified as follows

$$Q_{y_{t+h}}(\tau \mid F_t) = \omega_0(\tau) + \omega_1(\tau)\hat{y}_{t+h,t}^1 + \ldots + \omega_k(\tau)\hat{y}_{t+h,t}^k; \quad (7)$$

$$\omega_j(\tau) = (\omega_j + \gamma_j \gamma_h) \text{ and } j = 0, \ldots, k$$

$$\gamma_h = F^{-1}_{\eta,h}(\tau), \tau \in (0,1).$$

Equation (7) states that we can identify any quantile of $F_{t+h,t}$ through a quantile regression of $y_{t+h}$ on a constant and the vector of forecasts $\hat{y}_{t+h,t}$.
for various $\tau \in (0, 1)$. This result can be used to estimate the conditional density $f_{t+h,t}$ without imposing any parametric form on the forecast error distribution.

The forecasting literature has also suggested other approaches to combine probability density forecasts, which we briefly describe below.

### 2.1 Linear Combination of Density Forecasts

Stone (1961) considered the following linear combination

$$ f_{t+h,t}^c(y_{t+h}) = \sum_{j=1}^{k} \omega_{t+h,t,j} f_{t+h,t,j}(y_{t+h}) $$

(8)

where $f_{t+h,t,j}$ is the conditional density from the $j$th 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 maximizes the average log predictive score function as follows

$$ w^* = \arg \max_w 1 \frac{T}{T} \sum_{t=1}^{T} \ln \left( f_{t+h,t}^c(y_{t+h}) \right). $$

(9)

They show that when weights are chosen as in (9) then the combined density is optimal in the sense that it minimizes the following loss function

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

(10)

The optimality of the above combination of density functions depends on a known loss function (10) and full knowledge of the econometric models used to generate the individual densities. Our approach, on the other hand, relies only on individual point forecasts. This is particular important in situations where the decision maker only observes point forecasts from
economic institutions such as IMF, commercial banks, World Bank, and want to use them to construct optimal density forecasts without making assumptions on the parametric specification of the (unknown) models used by such institutions, which includes the specification on the forecast error distribution.

Furthermore, as discussed by Timmermann (2006), combining density forecast as in (8) imposes new requirements beyond those used for combination of point forecasts because there is a need to guarantee that probability forecasts will never become negative and will always sum to one. For this reason, one needs to impose the combination to be convex with weights confined to the zero-one interval and integrating to unity. Unfortunately there is no unique weighting scheme that satisfies such conditions. Indeed, Kascha and Ravazzolo (2010) evaluated combination (8) using different weighting schemes and concluded that equal weights and mean-square error weights provide more uniform results than the recursive log score weights suggested by Jore et al. (2010). They also suggested that different results can be obtained under different datasets and, therefore, there is no such optimal weights that one can use to calculate (8).\footnote{Another alternative approach based on the full knowledge of individual models is the one suggested by Granger (1969, 1989) which states that any conditional quantile of $y_{t+h}$ can be written as a linear combination of conditional quantiles estimated from different models, $q_{t+h,t,j}$, $j = 1,...,k$. Since this method is also based on the full knowledge of individual models, we do not include it in our empirical exercise.}

So far, we have presented our approach without a proper evaluation. In the next section, we use it to forecast the density of inflation in the United States and compare its results to the ones obtained from different methods currently employed in the literature.
3 Forecasting Inflation Densities

The main purpose of this section is to provide an empirical evidence to the theoretical results derived previously. 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 documents evidence in favor of the hypothesis that uncertainty on future inflation hurts economic growth. Uncertainty about inflation can also affect fiscal-policies-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 United States are affected by two major problems. Firstly, as documented by Clark (2011), the volatility of inflation in the U.S. decreased during the 1987-2007 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 has shown that
density forecasts are improved when one allows for the presence of time-varying variance (see Clark, 2011, and Jore, Mitchell and Vahey, 2010). A second concern is that the distribution function of the future inflation is probably unknown to 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. Second, estimation of the model is through quantile regression methods which does not require knowledge of $F_{y,h}$ and, therefore, addresses 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) were 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. Thus, combining forecasts across different models can be viewed as a way to make the forecast more robust against such a misspecification bias (Timmermann, 2006).

3.1 The Econometric Models

In this section, we will assume that there are 7 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 use the methods suggested by Hall and Mitchell (2007). Thus, the contribution of each individual model introduced in this section will be twofold.\(^5\) First, they will be employed to forecast the conditional mean, which is used as covariate in (7). Second, the models will be used to forecast the densities of inflation that will be combined by using the approach suggested by Hall and Mitchell (2007).

The first model is the Phillips curve, 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) and it can be cast in the following model

\[
\begin{align*}
y_{t+h} & = X'_{t+h,t} \alpha + \eta_{t+h} \\
y_{t+h} & = \pi_{t+h} - \pi_t \\
\eta_{t+h} & \sim iid \ Normal \ (0, \sigma^2).
\end{align*}
\]

where \(\pi_{t+h}\) is the inflation rate at some future time \(t+h\), \(\pi_t\) is the inflation rate at time \(t\), and \(X_{t+h,t}\) is a vector of economic indicators that are known at time \(t\). Model (11) admits a great variety of specifications and we will consider the following ones:

Model 1. Phillips Curve with market expectations: This is the model (11) with

\[
X'_{t+h,t} = (1, \pi_{t+h,t} - \pi_t, \Delta \pi_t, \Delta \pi_{t-1}, u_{t+h,t} - u_t, \Delta u_t) \quad \text{and} \quad \alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)',
\]

where \(\pi_{t+h,t}\) and \(u_{t+h,t}\) are respectively the market ex-

\(^5\)The term "model" is here used in a broad sense that includes forecasting methods.
expectations of inflation and unemployment rates (defined as the median forecasts from the survey of professional forecasters published by the FED of Philadelphia). Finally, $u_t$ is the unemployment rate at time $t$ and $\Delta$ is the first-difference operator.

Model 2. The same as 1 but with $\alpha_j = 0$ ; $j = 2, ..., 5$. In this specification, only the inflation market expectations matters to forecasting the future values of inflation.

Model 3. This is the model (11) but with $X'_{t+h,t} = \left(1, \sum_{j=0}^q \Delta \pi_{t-j}, \sum_{j=0}^k \Delta u_{t-j}\right)$ and the vector $\alpha$ is changed accordingly. In this specification, the market expectations on inflation and unemployment are replaced by adaptive ones $\sum_{j=0}^q \Delta \pi_{t-j}$ and $\sum_{j=0}^k \Delta u_{t-j}$, where $q$ and $k$ are chosen by the BIC information criterion.

Model 4. Same as 3, but with $\alpha_j = 0$ ; $j > q + 2$. In this specification, only the inflation adaptive expectations matters to forecasting the future values of inflation.

We also consider three additional models:

Model 5. It is a Gaussian GARCH(1,1) model with conditional mean given by:

$$\pi_{t+h} - \pi_t = \alpha_0 + \alpha_1 (\pi_{t+h,t} - \pi_t) + \alpha_2 \Delta \pi_t + \alpha_3 \Delta \pi_{t-1}. \quad (12)$$

Model 6. This is the so-called two-piece-normal-distribution model employed by the Bank of England to forecast inflation densities. According to Britton et al. (1998), the pdf for the so-called ‘two-piece’ normal distribution is given by $AS$, in which $A = \frac{2}{((1/\sqrt{1-\gamma})+(1/\sqrt{1+\gamma}))}$ and

---

6Section 4 presents a complete description of the data used to estimate each individual model.
\[ S = \frac{1}{\sqrt{2\pi}\sigma^2} e^{\left[ \frac{-(x-\mu)^2}{2\sigma^2} \right]} \]. In this paper, we construct a conditional density forecast based on this approach (model 6) and assuming that: (i) the conditional mean \( \mu \) is given by

\[
\pi_{t+h} - \pi_t = \alpha_0 + \alpha_1 (\pi_{t+h,t} - \pi_t) + \alpha_2 \Delta \pi_t + \alpha_3 \Delta \pi_{t-1} \\
+ \alpha_4 (u_{t+h,t} - u_t) + \alpha_5 \Delta u_t; \\
\]

(ii) the conditional variance \( \sigma^2 \) comes from a GARCH(1,1) estimation (distinct from the one of model 5) and; (iii) the conditional skew (\( \gamma \)), 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 \( \gamma = \frac{skw}{1+|skw|} \), which guarantees that \( \gamma \in [-1;1] \). Note that if the conditional skew is set to zero, then model 6 based density forecasts are the same as the ones obtained using a standard GARCH approach.

Model 7. It is a Gaussian AR(1) model with dependent variable given by \( y_{t+h} = \pi_{t+h} - \pi_t \).

Thus, models 1-4 and model 7 are constant-volatility models, model 5 has a time-varying volatility with symmetric distribution, and model 6 has a time-varying volatility with asymmetric distribution. Given the empirical evidence that the volatility of inflation is not constant over time, we expect that models 1-4 and 7 will not make good density forecasts. 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 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 using our approach, the decision maker does not need to care too much about the unrealistic specifications of models 1-7. Indeed, the advantage of our approach over models 1-7 is that: (i) we take advantage of the combination device to minimize the uncertainty about the correct specification of the conditional quantile function; (ii) no parametric specification is imposed on the error distribution function; (iii) it also allows for the presence of time-varying volatility and; (iv) unlike the combination methods (8), our approach does not require full knowledge of the econometric models, only observations on point forecasts are required.

In what follows, each of the above models will be used to estimate point forecasts $\hat{y}_{t+h,t}^j$ and densities $f_{t+h,t,j}$, $j = 1, 2, \ldots, 7$. Next, we combine $f_{t+h,t,j}$ by using (8) and use the point forecasts $\hat{y}_{t+h,t}^j$ to estimate the conditional quantiles of $y_{t+h}$ through our optimal combination approach (7). For the combination using (8), we follow Kascha and Ravazzolo (2010) and consider equal weights (labeled as model 8), MSE weights (labeled as model 9) and recursive log score weights (labeled as model 10). The optimal combination (7) will be labeled as model 11, respectively. Finally, for model 11, given a family of estimated conditional quantile functions, the conditional density of $y_{t+h}$ is estimated by the Epanechnikov kernel, which is a weighting function that determines the shape of the bumps. We prefer such a kernel because it generates smooth densities, especially when the time series sample size is short, which is the case in this empirical application of
inflation forecast. In the next section we describe the database as well as the forecasting schemes employed in this empirical exercise.

4 Data

Inflation is measured with GDP or GNP deflator, depending on data vintage, and the inflation rate is defined as the annualized log changes. We collect our data from the Federal Reserve Bank of Philadelphia’s Real Time Data Set for Macroeconomists (RTDSM). Our paper also uses data on quarterly market forecasts of unemployment and annualized inflation rates, measured as the respective median forecasts from the Survey of Professional Forecasters (SPF), published by the Federal Reserve Bank of Philadelphia. Since the SPF dataset begins in 1968Q4, for model estimation purposes we extended from 1968.Q3 back to 1961.Q1 the time series of unemployment and GDP price index expectations by using exponential smoothing.\footnote{Following Clark (2011), the exponentially smoothed series for the expected unemployment rate \( u_{t+1,t} \) is constructed as follows: (i) Initialize the filter with the average unemployment rate \( \langle u_t \rangle \) of 1948Q1-1959Q3. The average becomes the exponentially smoothed estimate for period 1959Q4; (ii) Use exponential smoothing formula \( u_{t+1,t} = \alpha u_t + (1 - \alpha) u_{t+1,t-1} \) with a (calibrated) smoothing parameter \( \alpha = 0.30 \) (inflation) and \( \alpha = 0.60 \) (unemployment) to estimate the trend unemployment expectation for 1960Q1 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 \( u_{t+h,t} = \alpha u_t + (1 - \alpha) u_{t+h-1,t} \).}

The starting point of the estimating sample is always 1961.Q1. In order to use the sample 1961Q1-1984Q4, we adopt vintage 1985Q1 (with information until 1984Q4). The only exception is vintage 1996Q1, due to data unavailability, in which information for GDP price index in 1995Q4 is obtained from 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 1961Q1 and ends at 1984Q4, 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 through 2008.Q4. For each forecast starting at the origin $t=1985.Q1$, we use the real-time data vintage $t$, which 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 the models 8-11, 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 1985Q1- $TS+1$ to 1985Q1 and from 1985Q1- $TS+1-h$ to 1985Q1-h for the individual model forecasts.

5 Results

We compute direct forecasts (Marcellino, Stock and Watson, 2006). At forecast horizons $h=1$ and $h=2$, the inflation rate being forecasted is the annualized quarterly rate, defined as $400 \log(P(t+h)/P(t+h-1))$, in which $P_t$ is the (real time) GDP (or GNP) price index. All other variables are also log-transformed. 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) the LR test of Berkowitz (2001), which evaluates the entire conditional density via probability integral transforms (PITs); (iii) log predictive density score, which allows one to rank the models; and (iv) 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
"...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 $t+h+2$ data set.

A natural starting point for forecast density evaluation is interval forecasts - that is, 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 1 reports the frequency with which actual real-time outcomes for inflation falls inside 70 percent intervals. Accurate intervals should result in frequencies of about 70 percent. A frequency of more (less) than 70 percent 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 70 percent based on t-statistics.

Table 1 shows the forecast coverage rate at the two horizons in both estimation schemes. In the recursive scheme, the models with constant volatility (models 1-4 and 7) generate densities (at $h = 1$ and 2) that are too wide with the null hypothesis of 70% coverage being easily rejected at a 5% level. The models with time-varying volatility (models 5 and 6) exhibit correct coverage at the recursive scheme but they generate narrow densities in the rolling-window estimation scheme. The combined density method (models 8, 9 and 10) suffers from the low performance of the individual models and also generate densities that are too wide. The combined density is sometimes even wider than the ones generated by the individual models. This
result has already been reported in the literature by Kascha and Ravazzolo (2010) who pointed out that the combination method (8) may generate densities that are generally more dispersed than the individual ones, which may result in a poor coverage rate. Based on the reported p-values, the density forecasted by our proposed combination method (model 11) has correct coverage at the two forecasting horizons and in both estimation schemes. In what follows we report similar results obtained by using the probability integral transforms (PITs) and log scores, which are a broader measure of density calibration.

Table 1 - Real-Time Forecast Coverage Rates
(frequencies of actual outcomes falling inside 70% interval band)

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>h=1</td>
<td>0.854</td>
<td>0.865</td>
<td>0.813</td>
<td>0.833</td>
<td>0.667</td>
<td>0.729</td>
<td>0.792</td>
<td>0.844</td>
<td>0.844</td>
<td>0.844</td>
<td>0.719</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(0.006)</td>
<td>(0.001)</td>
<td>(0.492)</td>
<td>(0.524)</td>
<td>(0.03)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.685)</td>
</tr>
<tr>
<td>h=2</td>
<td>0.875</td>
<td>0.875</td>
<td>0.917</td>
<td>0.885</td>
<td>0.750</td>
<td>0.760</td>
<td>0.896</td>
<td>0.896</td>
<td>0.906</td>
<td>0.896</td>
<td>0.771</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.302)</td>
<td>(0.218)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.203)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>h=1</td>
<td>0.802</td>
<td>0.802</td>
<td>0.792</td>
<td>0.760</td>
<td>0.604</td>
<td>0.583</td>
<td>0.750</td>
<td>0.792</td>
<td>0.792</td>
<td>0.792</td>
<td>0.760</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.01)</td>
<td>(0.171)</td>
<td>(0.059)</td>
<td>(0.023)</td>
<td>(0.263)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.171)</td>
</tr>
<tr>
<td>h=2</td>
<td>0.833</td>
<td>0.813</td>
<td>0.823</td>
<td>0.813</td>
<td>0.531</td>
<td>0.563</td>
<td>0.813</td>
<td>0.802</td>
<td>0.802</td>
<td>0.802</td>
<td>0.771</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.015)</td>
<td>(0.002)</td>
<td>(0.023)</td>
<td>(0.015)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.114)</td>
</tr>
</tbody>
</table>

Note: The table includes in parentheses p-values for the null of correct coverage

(empirical = nominal rate of 70%), based on t-statistics using standard errors

computed with the Newey-West estimator, with a bandwidth of 0 at the

1-quarter horizon and 1.5×horizon for the 2-quarter horizon.

We compute normal transforms of PITs which provide useful indicators of the calibration of density forecasts. The normalized forecast error is defined as $\Phi^{-1}(z_{t+1})$, where $z_{t+1}$ denotes the PIT of a one-step ahead forecast.

23
error and $\Phi^{-1}$ is the inverse of the standard normal distribution function. Berkowitz develops tests based on the normality of the normalized errors that have better power than tests based on the uniformity of the PITs. These tests have been used in recent studies such as Clements (2004), Jore, Mitchell and Vahey (2010), and Clark (2011). Table 2 reports the results of the Berkowitz test. For the recursive estimating scheme, the constant volatility models do not pass the Berkowitz test at a 5% level. Again, models 8, 9, and 10 suffer from the poor performance of the individual models and do not pass the test either. The results are a little better when we use the rolling-window scheme with the individual models as well as combined densities passing the Berkowitz test. The method proposed in this paper (model 11) again works very well and easily passes the Berkowitz test at both estimating schemes. We next rank each model based on their log-scores used in such recent studies as Geweke and Amisano (2010) and Clark (2011). Indeed, the overall calibration of the density forecasts can most broadly be measured with log predictive density scores, under which a higher score implies a better model.

<table>
<thead>
<tr>
<th>Table 2 - Berkowitz (2001) density test (p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recursive estimation</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>h=1</td>
</tr>
<tr>
<td>Rolling window</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>h=1</td>
</tr>
</tbody>
</table>

Table 3 reports the values of the log predictive density scores (LPDS). We easily notice that no other model has a LPDS that is higher than the one of model 11 at both estimating schemes and forecasting horizons. In general, the performance of the constant-variance models are very alike and
are outperformed by all other models. The combined densities (models 8, 9 and 10) outperform the constant volatility models, but have a LPDS that is smaller than the one from a time-varying model (models 5 and 6). As suggested by Kascha and Ravazzolo (2010), this result can be interpreted as an insurance against bad models provided by the linear combination (8). 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. Our approach is distribution free and the uncertainty about the specification of the quantile function is minimized through the combination device. 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 11 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 3 - Log predictive density score (LPDS)**

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
</table>

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

Table 4 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 11 and model $j$, $j \neq 11$ is rejected at a 5% level. Based on the p-values reported on Table 4, we can conclude that no other model has an LPDS that is statistically equal to the LPDS of model 11, reinforcing our previous results about the good performance of the proposed approach.

In sum, our empirical exercise indicates that our combination method is not outperformed by either any of the individual models used to generate the point forecasts in $\mathcal{F}_1$ nor by the combined densities obtained using the linear combination (8). 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.

**Table 4 - Amisano-Giacomini (2007) test applied to average LPDS**

<table>
<thead>
<tr>
<th>AG test: stat (p-value)</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
<th>Model 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h=1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.293</td>
<td>0.294</td>
<td>0.293</td>
<td>0.293</td>
<td>0.092</td>
<td>0.077</td>
<td>0.293</td>
<td>0.293</td>
<td>0.242</td>
<td>0.236</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.0135)</td>
<td>(0.017)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>$h=2$</td>
<td>0.509</td>
<td>0.509</td>
<td>0.509</td>
<td>0.509</td>
<td>0.260</td>
<td>0.271</td>
<td>0.509</td>
<td>0.509</td>
<td>0.455</td>
<td>0.444</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
</tbody>
</table>

Note: Null hypothesis of zero average difference in LPDS between model 11 (benchmark) and model $i = 1, \ldots, 10$. Similar to Clark (2011), the p-values 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 1-quarter horizon and 1.5×horizon in the 2-quarter horizon).
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 the 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, that is, 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 the 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 an 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 modelling 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 from both individual models and the combination method suggested by Hall and Mitchell (2007). The evidence presented in this paper shows that the proposed optimal combination method materially improves the real-time accuracy of density forecasts. The density evidence includes interval forecasts (coverage rates), tests applied to normal transforms of the probability integral transforms, 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), which was negatively affected in our exercise by the inclusion of constant-volatility models. 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 $\mathcal{F}_t$. In particular, we could follow the idea of Stock and Watson (1999) and expand $\mathcal{F}_t$ by including other Phillips curve models that are based on measures of economic activity other than the unemployment rate, or even consider Bayesian vector autoregressive models with stochastic volatilities as recently suggested by Clark (2011). In other words, the search for more models that could belong to $\mathcal{F}_t$ is still open and it is an interesting object for future research.
Appendix

**Proof of Proposition 1.** The proof is similar to the one 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:

\[
\hat{y}_{t+h,t} = \arg \min_{\hat{y}} \int L^i(y - \hat{y})dF_{t+h,t}(y) = \arg \min_{\hat{y}} \int \left[ g \left( \frac{1}{\hat{y}^i_{t+h,t}} \right) \right]^{-1} L^i \left( \frac{1}{\hat{y}^i_{t+h,t}} (y - \hat{y}) \right) dF_{t+h,t}(y)
\]

\[
= \arg \min_{\hat{y}} \int \left[ g \left( \frac{1}{\gamma_0 + \gamma_1 \hat{y}^i_{t+h,t} + \ldots + \gamma_k \hat{y}^i_{t+h,t}} \right) \right]^{-1} \cdot L^i \left( \frac{1}{\gamma_0 + \gamma_1 \hat{y}^i_{t+h,t} + \ldots + \gamma_k \hat{y}^i_{t+h,t}} (y - \hat{y}) \right) dF_{t+h,t}(y)
\]

\[
= \arg \min_{\hat{y}} \int L^i \left( \frac{1}{\gamma_0 + \gamma_1 \hat{y}^i_{t+h,t} + \ldots + \gamma_k \hat{y}^i_{t+h,t}} (y - \hat{y}) \right) dF_{t+h,t}(y) 
\]

\[
= \arg \min_{\hat{y}} \int L^i \left( \omega_0 + \omega_1 \hat{y}^i_{t+h,t} + \ldots + \omega_k \hat{y}^i_{t+h,t} + \gamma_0 \hat{y}^i_{t+h,t} + \gamma_1 \hat{y}^i_{t+h,t} \hat{y}^i_{t+h,t} + \ldots + \gamma_k \hat{y}^i_{t+h,t} \hat{y}^i_{t+h,t} - \hat{y} \right) dF_{t+h,t}(y).
\]

Let us represent a forecast by \(\omega_0 + \omega_1 \hat{y}^i_{t+h,t} + \ldots + \omega_k \hat{y}^i_{t+h,t} + \left( \gamma_0 + \gamma_1 \hat{y}^i_{t+h,t} + \ldots + \gamma_k \hat{y}^i_{t+h,t} \right) \hat{y}^i_{t+h,t}\). This way, it follows that:
\[
\hat{y}_{t+h,t} = \omega_0 + \omega_1 \hat{y}_{t+h,t}^1 + \ldots + \omega_k \hat{y}_{t+h,t}^k + \left( \gamma_0 + \gamma_1 \hat{y}_{t+h,t}^1 + \ldots + \gamma_k \hat{y}_{t+h,t}^k \right).
\]

where \( \omega_0 = \arg\min_{\hat{\gamma}} \int L_i \left( \frac{1}{\gamma_0 + \gamma_1 \hat{y}_{t+h,t}^1 + \ldots + \gamma_k \hat{y}_{t+h,t}^k} \right) \left( \omega_0 + \omega_1 \hat{y}_{t+h,t}^1 + \ldots + \omega_k \hat{y}_{t+h,t}^k + \left( \gamma_0 + \gamma_1 \hat{y}_{t+h,t}^1 + \ldots + \gamma_k \hat{y}_{t+h,t}^k \right) \right) dF_{\eta,h}(\eta)
\]

where \( \omega_0 = \arg\min_{\hat{\gamma}} \int L_i \left( \eta_{t+h} - \hat{\eta} \right) dF_{\eta,h}(\eta)
\]

\[
\omega_0(\tau_i) = (\omega_0 + \gamma_i \hat{\gamma}_h^i), \omega_j(\tau_i) = (\omega_j + \gamma_j \hat{\gamma}_h^j), j = 1, \ldots, k.
\]

\begin{proof}
If we assume that there are no scale effects then \( \gamma_1 = \ldots = \gamma_k = 0 \) therefore the optimal forecast will be \( \hat{y}_{t+h,t} = \omega_0(\tau_i) + \omega_1 \hat{y}_{t+h,t}^1 + \ldots + \omega_k \hat{y}_{t+h,t}^k \), where \( \omega_0(\tau_i) = (\omega_0 + \gamma_0 \hat{\gamma}_h^i) \). This proves corollary 2. To prove corollary 1 we remember that the expected value of the MSE loss is \( \hat{y}_{t+h,t} = \arg\min_{\hat{y}} E(y - \hat{y})^2 \). Now, due to certainty equivalence \( E(y - \hat{y})^2 \) is minimized at \( \hat{y} = E(y|\hat{y}) = \omega_0 + \omega_1 \hat{y}_{t+h,t}^1 + \ldots + \omega_k \hat{y}_{t+h,t}^k \). This proves corollary 1.
\end{proof}

References


