ISSN 1518-3548

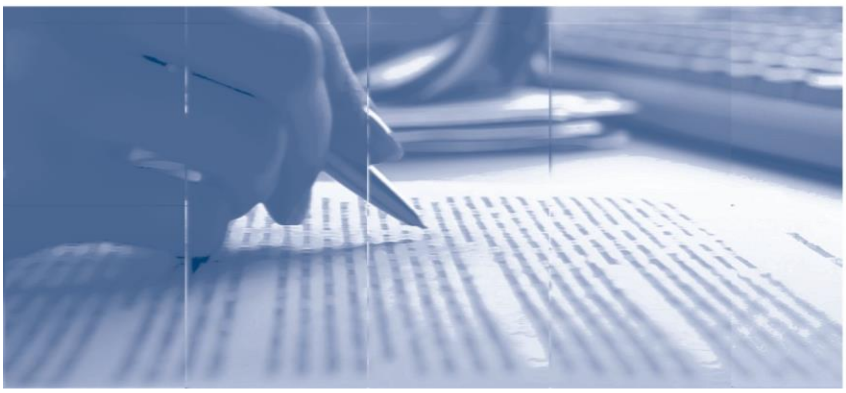
BANCO CENTRAL DO BRASIL

Microfounded Forecasting

Wagner Piazza Gaglianone and
João Victor Issler

December, 2014

Working Papers



372

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	n. 372	December	2014	p. 1-48

Working Paper Series

Edited by Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Francisco Marcos Rodrigues Figueiredo – E-mail: francisco-marcos.figueiredo@bcb.gov.br
Editorial Assistant: Jane Sofia Moita – E-mail: jane.sofia@bcb.gov.br
Head of Research Department: Eduardo José Araújo Lima – E-mail: eduardo.lima@bcb.gov.br

The Banco Central do Brasil Working Papers are all evaluated in double blind referee process.

Authorized by Carlos Hamilton Vasconcelos Araújo, Deputy Governor for Economic Policy.

General Control of Publications

Banco Central do Brasil

Comun/Dipiv/Coivi

SBS – Quadra 3 – Bloco B – Edifício-Sede – 14º andar

Caixa Postal 8.670

70074-900 Brasília – DF – Brazil

Phones: +55 (61) 3414-3710 and 3414-3565

Fax: +55 (61) 3414-1898

E-mail: editor@bcb.gov.br

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: <http//www.bcb.gov.br/?CONTACTUS>

Microfounded Forecasting*

Wagner Piazza Gaglianone†

João Victor Issler‡

Abstract

In this paper, we propose a microfounded framework to investigate a panel of forecasts (e.g. model-driven or survey-based) and the possibility to improve their out-of-sample forecast performance by employing a bias-correction device. Following Patton and Timmermann (2007), we theoretically justify the modeling of forecasts as function of the conditional expectation, based on the optimization problem of individual forecasters. This approach allows us to relax the standard assumption of mean squared error (MSE) loss function and, thus, to obtain optimal forecasts under more general functions. However, different from these authors, we apply our results to a panel of forecasts, in order to construct an optimal (combined) forecast. In this sense, a feasible GMM estimator is proposed to aggregate the information content of each individual forecast and optimally recover the conditional expectation. Our setup can be viewed as a generalization of the three-way forecast error decomposition of Davies and Lahiri (1995); and as an extension of the bias-corrected average forecast of Issler and Lima (2009). A real-time forecasting exercise using Brazilian survey data illustrates the proposed methodology.

Keywords: Forecast Combination, Common Features, Panel Data.

JEL Classification: C14, C33, E37.

*We are especially grateful for the comments and suggestions given by Antonio Galvão, Fabian Krüger, Marcelo Moreira, Mark Watson, Rafael Santos, Raffaella Giacomini and Scott Atkinson. We also benefited from comments given by the seminar participants of the 33rd International Symposium on Forecasting (Seoul); 19th International Panel Data Conference (London); 2013 Latin American Workshop in Econometrics (São Paulo); XVI Annual Inflation Targeting Seminar of the Banco Central do Brasil (Rio de Janeiro); and 1st Conference of the International Association for Applied Econometrics (London). Any remaining errors are ours.

†Research Department, Banco Central do Brasil. E-mail: wagner.gaglianone@bcb.gov.br

‡Corresponding Author. Graduate School of Economics, Getulio Vargas Foundation, Praia de Botafogo 190, s.1104, Rio de Janeiro, RJ 22.253-900, Brazil. E-mail: jissler@fgv.br

1 Introduction

Constructing accurate forecasts of macroeconomic variables is crucial for policymakers, firms and consumers. In particular, reliable inflation forecasts are of utmost importance for monetary policy and it is no surprise that a considerable academic literature investigates a vast amount of inflation forecasts and forecasting methods.

According to Ang et al. (2007), economists use four main methods to forecast inflation. The first method is atheoretical, using time series models of the ARIMA variety. The second method builds on the economic model of the Phillips curve, leading to forecasting regressions that use real activity measures. Third, forecast inflation can be constructed by using information embedded in asset prices, in particular the term structure of interest rates. Finally, survey-based measures use information from agents (consumers or professionals) directly to forecast inflation. In this sense, the authors find that true out-of-sample survey forecasts (e.g. Michigan; Livingston) outperform a large number of out-of-sample multivariate time series competitors.

In the same line, Faust and Wright (2012) argue that subjective forecasts of inflation seem to outperform model-based forecasts in certain dimensions, often by a wide margin. The authors discuss some reasons why the subjective forecasters outperform most conventional models, pointing out to the choice of boundary values and the fact that professional forecasters quite often have access to econometric models and add expert judgment to these models.

The purpose of this paper is to propose a microfounded framework to investigate a given set of inflation forecasts (e.g. model-driven or survey-based) and the possibility to improve their out-of-sample forecast performance by employing an average forecast bias-correction device. The combination of forecasts is long used to diversify out the risk of large forecast errors.[1] This approach, first put forward by Bates and Granger (1969), has been shown to reduce forecast uncertainty of individual models

[1] Regardless of how one combine forecasts, if the series being forecasted is stationary and ergodic, and there is enough diversification among forecasts, one should expect that a weak law-of-large-numbers (WLLN) applies to well-behaved forecast combinations. The reason why forecast combination, in general, works well is because it takes advantage of the principle of risk diversification: idiosyncratic forecast errors vanish because of the WLLN works as the number of forecasts being combined increases without bounds. Timmermann (2006) uses such a risk diversification argument to defend pooling of forecasts.

in a variety of studies (see Hendry and Clements, 2002; Stock and Watson, 2006; Capistrán and Timmermann, 2009).

In particular, Granger and Ramanathan (1984) investigate optimal forecast combination under mean squared error (MSE) loss. The authors propose to estimate optimal weights from an ordinary least squares (OLS) regression of the variable of interest on a vector of forecasts plus an intercept to account for model (additive) bias. Palm and Zellner (1992) discuss whether "to pool or not to pool" forecasts using a two-way decomposition of the forecast error. Davies and Lahiri (1995) develop a three-way decomposition, although focusing on rationality instead of forecast combination (see also Lahiri et al., 2013). More recently, Issler and Lima (2009) propose an optimal forecast-combination, where forecasts of different models or survey results comprise the cross-sectional dimension in a panel data approach. In their context, the optimal forecast using a MSE risk function can be consistently estimated using a bias-corrected average forecast (BCAF).

Regardless of the forecast combination scheme, the focus of the existing literature often relies on the identification of the conditional expectation of the series being forecasted. If y_t is the series of interest and h is the forecast horizon, then, what is to be identified is the latent variable $\mathbb{E}_{t-h}(y_t)$, where $\mathbb{E}_{t-h}(\cdot)$ is the conditional expectation operator using all information available (observable or not) up to period $t-h$. In other words, the information content of every individual model (or forecast) is aggregated to optimally recover the conditional expectation.

In this paper, we tackle this issue by using a microfounded setup. We theoretically justify the modeling of individual forecast as a function of the conditional expectation, based on the individual optimization problem of forecasters, which, in turn, depends on the individual loss function and the data-generating process (DGP) of y_t. Here, we follow the work of Patton and Timmermann (2007), in the sense of investigating forecast optimality under more general loss functions (other than the usual MSE). However, different from these authors, we apply our results to a whole set of forecasts, in order to construct an optimal forecast combination scheme. In our setup, we assume that each individual forecaster provides a signal about $\mathbb{E}_{t-h}(y_t)$, and we use these signals, under mild conditions, to generate an optimal (combined) forecast.

This way, our approach allows us to derive a signal-extraction model from "first principles", instead of (*ad hoc*) assuming a relationship between the individual forecast and the conditional expectation. Furthermore, our setup enables us to extend the standard signal-extraction approach focused only on additive bias (e.g. Issler and Lima, 2009). Our contribution is, thus, to investigate different cases regarding the loss and DGP, and to reveal the resulting relationship between the optimal forecast and the conditional expectation in such cases (e.g. in the presence of an asymmetric loss function). In this paper, we also discuss model identification of the extended approach and propose consistent estimators for the suggested model. As a result, we provide tools to better understand the dynamics of survey-based forecasts and the related forecast revision process.

The rest of the paper is divided as follows. Section 2 presents a discussion on bias-correction devices and introduces a microfounded-based framework to study the forecast error decomposition under risk functions more general than the usual MSE. Section 3 presents a real-time forecasting exercise with data from a survey of Brazilian inflation expectations using the methods proposed here, comparing the (pseudo) out-of-sample performance of different bias-correction approaches. Section 4 concludes.

2 Econometric Setup

2.1 The bias-corrected average forecast (BCAF)

There are many potential sources of forecast bias available in the literature, besides the traditional argument of model misspecification (or parameter uncertainty). For instance, forecasters may have economic incentives (e.g. strategic behavior, competition or reputational concerns) to make biased forecasts (Laster et al., 1999; Ottaviani and Sorensen, 2006; Batchelor, 2007). In turn, biased forecasts can arise under asymmetric loss functions over forecast errors (Elliott et al., 2008; Capistrán and Timmermann, 2009). On the other hand, informational rigidities can play an important role regarding the behavior of forecasters (Mankiw and Reis, 2010; Coibion and Gorodnichenko, 2012).

In this paper, we follow the "asymmetric loss" approach in order to better understand the source of a time-invariant bias.[2] The techniques discussed in this section are appropriate for forecasting a weakly stationary and ergodic univariate process $\{y_t\}$ using a large number of forecasts that will be combined to yield an optimal forecast in the mean-squared error (MSE) sense (in next section, we relax such assumption allowing for a more general class of loss functions).

These forecasts could be the result of using several econometric models that need to be estimated prior to forecasting, or the result of using no formal econometric model, for instance, as the result of an opinion poll on the variable in question using a large number of individual responses. We can also imagine that some (or all) of these poll responses are generated using econometric models, but then the econometrician that observes these forecasts has no knowledge of them (e.g. survey of forecasts).

We label forecasts of y_t, computed using conditioning information sets lagged h periods, by $f_{i,t}^h$, $i = 1, 2, \ldots, N$. Therefore, $f_{i,t}^h$ are h-step-ahead forecasts of y_t, formed at period $(t-h)$, and N is either the number of models estimated to forecast y_t or the number of respondents of an opinion poll regarding y_t.

Forecasts $f_{i,t}^h$ are initially assumed to be approximations to the optimal forecast $\mathbb{E}_{t-h}(y_t)$ as follows:[3]

$$f_{i,t}^h = \mathbb{E}_{t-h}(y_t) + k_i^h + \varepsilon_{i,t}^h, \tag{1}$$

where k_i^h is the individual model time-invariant bias for h-step-ahead prediction and $\varepsilon_{i,t}^h$ is the individual model error term in approximating $\mathbb{E}_{t-h}(y_t)$, where $\mathbb{E}\left(\varepsilon_{i,t}^h\right) = 0$ for all i, t and a given h. In addition, one can always decompose the series y_t into $\mathbb{E}_{t-h}(y_t)$ and an unforecastable component ζ_t^h, such that $\mathbb{E}_{t-h}\left(\zeta_t^h\right) = 0$, as it follows:

$$y_t = \mathbb{E}_{t-h}(y_t) + \zeta_t^h. \tag{2}$$

Combining (1) and (2) yields the well known two-way decomposition, or error-

[2] Although there might be evidence of time-varying bias in some surveys, we assume along this paper that forecast bias is time-invariant.

[3] Recall that the conditional mean is an optimal forecast under a particular loss function (MSE) and some mild conditions on the data-generating process of y_t. More general loss functions will be considered in next section, in order to explain the forecast bias as the outcome of the individual optimization problem.

component decomposition, of the forecast error $f_{i,t}^h - y_t$ (Wallace and Hussain, 1969; Fuller and Battese, 1974):

$$\begin{aligned} f_{i,t}^h - y_t &= \mu_{i,t}^h \qquad i = 1, 2, \ldots, N \\ \mu_{i,t}^h &= k_i^h + \eta_t^h + \varepsilon_{i,t}^h, \text{ where } \eta_t^h = -\zeta_t^h. \end{aligned} \tag{3}$$

Notice that by construction, the framework in (3) specifies explicit sources of forecast errors (see Issler and Lima, 2009) that are found in both y_t and $f_{i,t}^h$. The term k_i^h is the time-invariant forecast bias of model i (or of respondent i). It captures the long-run effect of forecast-bias of model i, or, in the case of surveys, the time invariant bias introduced by respondent i. Its source is $f_{i,t}^h$. The term η_t^h arises because forecasters do not have future information on y between $t-h+1$ and t. Hence, the source of η_t^h is y_t, and it is an additive aggregate zero-mean shock affecting equally all forecasts. The term $\varepsilon_{i,t}^h$ captures all the remaining errors affecting forecasts, such as those of idiosyncratic nature and others that affect some but not all the forecasts (a group effect).

From the perspective of combining forecasts, the components k_i^h, $\varepsilon_{i,t}^h$ and η_t^h play very different roles. If we regard the problem of forecast combination as one aimed at diversifying risk, i.e., a finance approach, then, on the one hand, the risk associated with $\varepsilon_{i,t}^h$ can be diversified, while that associated with η_t^h cannot. On the other hand, in principle, diversifying the risk associated with k_i^h can only be achieved if a bias-correction term is introduced in the forecast combination, which reinforces its usefulness.

Based on the previous two-way decomposition of the forecast error, Issler and Lima (2009) propose non-parametric consistent estimates for the components k_i^h, η_t^h and $\varepsilon_{i,t}^h$. They show that (under a set of mild conditions) the feasible bias-corrected average forecast (BCAF) given by $\frac{1}{N}\sum\limits_{i=1}^{N} f_{i,t}^h - \widehat{B^h}$, where $\widehat{B^h} = \frac{1}{N}\sum_{i=1}^{N} \widehat{k_i^h}$, obeys:

$$\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}} \left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h} \right) = y_t + \eta_t^h = \mathbb{E}_{t-h}(y_t), \tag{4}$$

where $\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}}$ is the probability limit using the sequential asymptotic setup of

Phillips and Moon (1999). The authors show that feasible BCAF is an optimal forecasting device under a mean-squared error (MSE) risk function. Notice that bias correction (in this setup) is essentially a form of intercept correction. Intuitively, if $k_i^h > 0$, model i will consistently over predict the target variable y_t and it is reasonable to correct its forecasts downwards by the same amount as k_i^h.

2.2 Under a more general risk function

In this section, we theoretically justify the presence of a slope coefficient associated to the conditional expectation in model (1), due to the presence of more general risk functions, such as asymmetric ones, as it follows:

$$f_{i,t}^h = \beta_i^h \mathbb{E}_{t-h}(y_t) + k_i^h + \varepsilon_{i,t}^h. \tag{5}$$

To do so, we investigate the optimization problem solved by individual forecasters, and analyze several assumptions for both the loss function and the data-generating process (DGP) of y_t, which might result in the previous equation. Indeed, departures from the standard MSE loss have been investigated in previous studies. For instance, in the context of forecast combination, Elliott and Timmermann (2004) argue that it is the combination of asymmetry in both the loss function and DGP that is required for optimal weights to differ from the MSE weights.

The novelty in this paper is to study these potential asymmetries in order to improve forecast performance by using a feasible bias-correction device applied to a whole set of forecast horizons. In this sense, we next investigate the relationship between the optimal forecast and the conditional expectation, based on the following cases: (1) Loss function and DGP both known by individual i (analytical relationship, in which DGP is assumed to be fully parametric or following a location-scale model); (2) Loss known but parameters of the DGP are unknown by individual i (i.e. estimated relationship); and (3) Loss and DGP unknown (GMM estimation by the econometrician, who only observes y_t and $f_{i,t}^h$).

Case 1: Loss and DGP known

Firstly, assume that there is an amount of individuals i forecasting y_t conditional on $\mathcal{F}_{t-h}$. Each individual chooses an optimal forecast $\widetilde{f}_{i,t}^h$ by minimizing its respective expected loss function L^i.

Assumption A1 (Loss function) L^i depends solely[4] on the forecast error $e_{t,t-h}^i \equiv y_t - \widetilde{f}_{i,t}^h$, that is, $L^i = L(e_{t,t-h}^i)$

Following Granger and Newbold (1986), it is a fairly natural criterion, given a loss function, to choose the forecast $\widetilde{f}_{i,t}^h$ so that the conditional expected loss is minimized. In this sense, define the optimal (point) individual forecast of y_t as:

$$\widetilde{f}_{i,t}^h \equiv \arg\min_f \ \mathbb{E}\left(L^i(y_t; f) \mid \mathcal{F}_{t-h}\right) \tag{6}$$

where $f \in \mathbb{R}$ are all possible choices of the $i-th$ forecaster and $\mathbb{E}(. \mid \mathcal{F}_{t-h})$ denotes the conditional expectation given $\mathcal{F}_{t-h}$. A natural assumption about the shape of the agent's loss function, which is unknown to the econometrician, is that if one forecasts without error, then no forecast cost arises, but if there is an error, then the larger it is the greater will be the loss function value, as it follows:

Assumption A2 (Loss function - shape) The loss function exhibits the following properties: (i) $L^i(0) = 0$; (ii) $L^i(e_i)$ is continuous, homogeneous and non-negative $\forall e_i \in \mathbb{R}$; and (iii) $L^i(e_i)$ is monotonic non-decreasing (for $e_i > 0$ or $e_i < 0$), and differentiable at least twice almost everywhere.

In practical terms, the symmetry of the loss function might be a restrictive hypothesis to be considered by an econometrician. Granger and Newbold (1986, p.125) provide two examples of situations where nonsymmetrical cost functions arise. In these cases, it would be interesting to check if the agent forecast is optimal under a broader class of loss functions. A simple way to consider an asymmetric function, and account for some "degree of asymmetry", is given by the following assumption:

[4] This is the same Assumption L1 of Patton and Timmermann (2007). According to the authors, although it rules out certain loss functions (e.g., those which also depend on the level of the predicted variable), many common loss functions are of this form.

Assumption A3 (Loss function - asymmetry) The loss function $L^i(e_i)$ can be decomposed as $L^i(e_i) = g^i(e_i)h^i(e_i)$, where $g^i(e_i)$ is a non-negative and symmetric function about $e_i = 0$; $g^{i\prime}(e_i)$ and $g^{i\prime\prime}(e_i)$ exist almost everywhere; $h^i(e_i) = \begin{cases} \beta_1^i & ; \ e_i < 0 \\ \beta_2^i & ; \ e_i > 0 \end{cases}$ where $\{\beta_1^i; \beta_2^i\}$ are positive constants.

Assumption A4 (DGP - stationarity and regularity of cdf) The univariate time series y_t is a weakly stationary and ergodic process and the conditional cumulative distribution functions (cdf) of y_t given $\mathcal{F}_{t-h}$ (denoted by $F_{t,t-h}(\cdot)$ or $F_t(\cdot \mid \mathcal{F}_{t-h})$) are absolutely continuous, with continuous densities $f_{t,t-h}$ uniformly bounded away from 0 and ∞ at the points $F_{t,t-h}^{-1}(\tau)$, $\forall \tau \in (0;1)$.

The additional assumptions A2 and A3 are made to investigate the shape of the loss function and are usually adopted in the literature (e.g. Granger and Newbold, 1986; Patton and Timmermann, 2007). Note that A3 is a simple generalization of the symmetric case, in which $\beta_1 = \beta_2$. A3 is quite general, covering a great deal of loss functions commonly mentioned in the literature, such as: mean squared error (MSE), mean absolute error (MAE), asymmetric linear (Lin-Lin), asymmetric quadratic, among many others. A4 is just a technical assumption about the DGP of y_t.

Proposition 1 *(Asymmetric Loss) If A1-A4 hold, then: (i) If $\beta_1 \neq \beta_2$ then $F_{t,t-h}(\widetilde{f}_{i,t}^h) \neq 0.5$, where $F_{t,t-h}$ is the conditional cdf of y_t; (ii) If $\beta_1 > \beta_2$ then $\widetilde{f}_{i,t}^h < Med_{t-h}(y_t)$; (iii) If $\beta_1 < \beta_2$ then $\widetilde{f}_{i,t}^h > Med_{t-h}(y_t)$, where $Med_{t-h}(y_t)$ is the conditional median of y_t; and (iv) for two forecasters i and j such that $\beta_1^i/\beta_2^i < \beta_1^j/\beta_2^j < 1$, then, $\widetilde{f}_{i,t}^h > \widetilde{f}_{j,t}^h > Med_{t-h}(y_t)$.*

Proposition 1 states two important results: (i) asymmetric loss functions generate departures of the optimal forecast from the central tendency (e.g. median) of the conditional distribution of y_t; and (ii) for the same DGP, the higher the degree of asymmetry in the loss function the greater will be the distance between the optimal forecast and the conditional median.

Corollary 2 *(Symmetric Loss) If A1-A4 hold, and $\beta_1 = \beta_2$, then $\widetilde{f}_{i,t}^h = Med_{t-h}(y_t)$.*

In other words, if $\beta_1 = \beta_2$, then, the optimal forecast is equal to the conditional median of y_t. Moreover, by considering additional assumptions on the DGP of y_t, we obtain the well-known optimality result of the conditional mean, due to Granger (1969, Theorem 2).

Corollary 3 *(Granger (1969): Optimality of the Conditional Mean) If A1-A4 hold and the conditional density function (pdf) of y_t is a unimodal, continuous and symmetric function about its conditional mean $\mathbb{E}_{t-h}(y_t) = \int_{-\infty}^{\infty} y_t f_{t,t-h}(y)dy$, and if $\beta_1 = \beta_2$ then $\widetilde{f}_{i,t}^h = \mathbb{E}_{t-h}(y_t)$.*

Note that an optimal forecast obtained, for instance, from a symmetric loss function such as MSE implies that $\widetilde{f}_{i,t}^h = \mathbb{E}_{t-h}(y_t)$. In fact, the classical result of Granger (1969), in which the optimal forecast equals the conditional mean of y_t under a MSE loss function, can be viewed as a special case of Proposition 1.

The exact relationship between the optimal forecast's percentile level τ_i, in respect to the cdf of y_t, and the parameters $\{\beta_1; \beta_2\}$ can be obtained if one considers a more restrictive assumption on the class of loss functions. The next corollary presents an example.

Corollary 4 *(Granger and Newbold (1986): Linear loss function) If A1-A4 hold and the loss function is the Lin-Lin function, i.e.,* $L(e_i) = \begin{cases} -\beta_1 e_i & ; \ e_i < 0 \\ 0 & ; \ e_i = 0 \\ \beta_2 e_i & ; \ e_i > 0 \end{cases}$, *in which $\{\beta_1; \beta_2\} > 0$, then, $\widetilde{f}_{i,t}^h = F_{t,t-h}^{-1}(\tau_i)$ in which $\tau_i = \beta_2/(\beta_1 + \beta_2)$.*

Under a symmetric loss function such as MSE, the only forecast that is unbiased[5] is the optimal forecast given by $\widetilde{f}_{i,t}^h = \mathbb{E}_{t-h}(y_t)$. However, under an asymmetric loss function, it is natural to expect that the optimal forecast will differ from the respective conditional mean. Intuitively, an asymmetric loss with, say, $\beta_1 > \beta_2$ indicates that the negative forecast errors are more costly to the forecaster than the positive ones. Thus, an individual forecaster will choose an optimal forecast that corresponds to some low quantile of y_t (i.e., $\tau_i < 0.5$) and therefore it is quite

[5] See Granger and Newbold (1986, p.144).

natural to expect that positive errors are more likely to be observed in historical data, which explains the forecast bias.

Granger (1969) argues that symmetry of both the loss function and the conditional density of y_t is not a sufficient condition for the optimum predictor to be equal to the conditional mean. In fact, the author provides a counter-example in which the conditional mean would be sub-optimal under symmetric functions (both loss and the pdf). In order to better understand the relationship between the optimal forecast and the conditional mean, more assumptions on the DGP and/or the loss function are required. In this sense, next Proposition analytically reveals, for illustrative purposes, this relationship for selected parametric forms of the pdf of y_t.

Proposition 5 *(DGP - parametric pdfs) If A1-A4 hold and the conditional pdf of* y_t *is: (i) Gaussian, Two-piece normal, or Logistic, then,* $\widetilde{f}_{i,t}^h = k_i^h + \mathbb{E}_{t-h}(y_t)$*; (ii) Log-normal or Weibull, then,* $\widetilde{f}_{i,t}^h = \beta_i^h \mathbb{E}_{t-h}(y_t)$*; (iii) Beta(a=1,b>0), then,* $\widetilde{f}_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t)$*; (iv) Beta(a>0,b=1), then,* $\widetilde{f}_{i,t}^h = \beta_i^h \xi(\mathbb{E}_{t-h}(y_t); \tau_i)$*, in which* ξ *is the non-linear function* $\exp(\frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)})$ *and* $\tau_i \equiv F_{t,t-h}(\widetilde{f}_{i,t}^h)$*.*

In a more realistic route, Proposition 6 assumes a location-scale model for the DGP of y_t and shows that, in this case, the optimal forecast is a linear function (i.e. intercept and slope) of the conditional mean. This class of DGPs is very convenient since it can naturally be investigated through the lens of a quantile regression framework with linear conditional quantiles.[6]

Assumption A5 (DGP - location-scale) The DGP of y_t follows a location-scale model, with conditional mean and variance dynamics defined as $y_t = X_{t,t-h}'\delta + \left(X_{t,t-h}'\gamma\right)\eta_t$, in which $(\eta_t|\mathcal{F}_{t-h}) \sim i.i.d.\ F_{\eta,h}(0,1)$, where $F_{\eta,h}(0,1)$ is some distribution with zero mean and unit variance, which depends on h but does not depend on $\mathcal{F}_{t-h}$; $X_{t,t-h} \in \mathcal{F}_{t-h}$ is a $m \times 1$ vector of covariates (which includes the intercept, and that can be predicted using information available at time $t-h$) and $\delta = [\delta_0; \delta_1; ...; \delta_{m-1}]$ and $\gamma = [\gamma_0; \gamma_1; ...; \gamma_{m-1}]$ are $m \times 1$ vectors of parameters. Without loss of generality, assume that $X_{t,t-h}' = (1, x_{t,t-h})$ is a 2×1 vector and $\delta = (\delta_0, \delta_1)'$; $\gamma = (\gamma_0, \gamma_1)'$.

[6] The linear quantile regression setup could be further extended to consider models with dependency and mixing conditions (e.g. Cai and Xiao, 2012; Galvao and Wang, 2013). We leave this route as a suggestion for future research.

The class of DGPs covered by A5 is very broad and includes most common volatility processes with time-varying variance (e.g. ARCH and stochastic volatility). Notice that no parametric structure is placed on $F_{\eta,h}$ and the covariate affects both the location and the scale of the conditional distribution of y_t.

Moreover, A5 implies that: (i) $Q_\tau(y_t \mid \mathcal{F}_{t-h}) = \alpha_0(\tau) + \alpha_1(\tau)x_{t,t-h}$ for some $\tau \in [0,1]$; and (ii) $\mathbb{E}(y_t \mid \mathcal{F}_{t-h}) = \mathbb{E}_{t-h}(y_t) = \overline{\alpha_0} + \overline{\alpha_1}x_{t,t-h}$; where $Q_\tau(.)$ is the conditional quantile of y_t, $[\alpha_0(\tau); \alpha_1(\tau)]$ depends on $(\delta;\gamma)$, L^i and $F_{\eta,h}(0,1)$; and $\overline{\alpha_j} \equiv \int_0^1 \alpha_j(\tau)d\tau$ for $j = \{0;1\}$. The previous expressions for both the conditional quantiles and the conditional expectation of y_t (under A5) will be next explored to deliver a linear connection between the optimal forecast and the conditional mean.

Proposition 6 *(Location-scale model) If A1-A5 hold, then: (i) the optimal forecast is a linear function of the conditional mean of* y_t*, so that* $\widetilde{f}_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t)$*; (ii) in the absence of scale effects on the DGP* ($\gamma_1 = \gamma_2 = ... = \gamma_{m-1} = 0$) *it follows that* $\beta_i^h = 1$ *and, thus,* $\widetilde{f}_{i,t}^h = k_i^h + \mathbb{E}_{t-h}(y_t)$.

Case 2: Loss known but parameters of DGP unknown

So far, we have assumed that the DGP of y_t is known to the individual i forecasting y_t, although (in practice) only its own loss function would be known. In other words, an optimal forecast of the form $\widetilde{f}_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t)$ would (in practice) be not feasible. Nonetheless, individuals might approximate the optimal forecast by using its finite sample (and feasible) counterpart $f_{i,t}^h = \widehat{k}_i^h + \widehat{\beta}_i^h \widehat{\mathbb{E}}_{t-h}(y_t)$, which can be estimated by using available data. This way, an optimal forecast "approximation error" $\varepsilon_{i,t}^h \equiv f_{i,t}^h - \widetilde{f}_{i,t}^h$ might arise, as illustrated in some examples presented in the Appendix. Now, define $\widehat{B^h} \equiv \frac{1}{N}\sum_{i=1}^{N} \widehat{k}_i^h$; $\widehat{\beta}^h \equiv \frac{1}{N}\sum_{i=1}^{N} \widehat{\beta}_i^h$ and consider the additional assumption:

Assumption A6 $[\widehat{k}_i^h; \widehat{\beta}_i^h] \equiv [\widehat{\alpha}_0(\tau_i) - \frac{\widehat{\overline{\alpha_0}}}{\widehat{\overline{\alpha_1}}}\widehat{\alpha}_1(\tau_i); \frac{\widehat{\alpha}_1(\tau_i)}{\widehat{\overline{\alpha_1}}}]$, where $\widehat{\alpha}(\tau_i) \equiv [\widehat{\alpha_0}(\tau_i); \widehat{\alpha_1}(\tau_i)]$ are standard estimators for the linear quantile regression of y_t onto $[1; x_{t,t-h}]$, and $\widehat{\overline{\alpha_j}}$ is computed over a discrete grid of equidistant quantiles $\tau_i \in [\tau_1, \tau_2, ..., \tau_K]$, such that $\widehat{\overline{\alpha_j}} \equiv \frac{1}{K}\sum_{k=1}^{K} \widehat{\alpha}_j(\tau_k)\Delta\tau_k$, for $j = \{0;1\}$.

Proposition 7 *If A1-A6 hold then: (i) the optimal (feasible) forecast of y_t conditioned on $\mathcal{F}_{t-h}$ is of the form: $f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h$, where $\varepsilon_{i,t}^h$ accounts for finite sample parameter uncertainty; (ii) $[\widehat{k}_i^h; \widehat{\beta}_i^h]$ are consistent estimators for $[k_i^h; \beta_i^h]$; (iii) the Extended BCAF (Bias Corrected Average Forecast), given by $\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}}$, asymptotically converges to the conditional mean of y_t: $\underset{(T,N\to\infty)_{\text{seq}}}{plim} \left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}} \right) = \mathbb{E}_{t-h}(y_t)$.*

Therefore, we showed that the relationship between the optimal forecast and the conditional mean is linear (intercept and slope) in several cases. It is worth mentioning that standard linear models, for instance, used for inflation forecasting (e.g. Phillips curve, autoregression, stochastic volatility or some factor models) can be nested into the previous location-scale specification (Stock and Watson, 1999; Faust and Wright, 2012).

For different DGPs (other than the previously considered), one can assume that: (i) the optimal forecast $\widetilde{f}_{i,t}^h$ is a linear combination of the conditional mean, such that $\widetilde{f}_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t)$; and (ii) the observed forecast $f_{i,t}^h$ can be viewed as a linear approximation of the individual optimal forecast $\widetilde{f}_{i,t}^h$, such that the following equation holds:

$$f_{i,t}^h = \widetilde{f}_{i,t}^h + \varepsilon_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h \tag{7}$$

or, equivalently,

$$f_{i,t}^h = k_i^h + \beta_i^h (y_t + \eta_t^h) + \varepsilon_{i,t}^h. \tag{8}$$

More general setups of the loss and the DGP of y_t are (of course) possible, although leading to (potential) non-linear relationships between the optimal predictor and the respective conditional mean, possibly embodied with non-tractable expressions of difficult practical use. In next section, we discuss model identification and the joint estimation of coefficients k_i^h and β_i^h within a generalized method of moments (GMM) setup.

Case 3: Loss and DGP unknown

Now, consider that an econometrician only observes a survey of individual forecasts $f_{i,t}^{h}$ and the target variable y_t, but has no information at all about the DGP and the individual loss functions. In this case, a panel framework can be employed to estimate the model coefficients and construct our extended bias-corrected average forecast, which is later showed to be optimal under the MSE loss function. The econometric model is the following:

$$f_{i,t}^{h} = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^{h} \tag{9}$$

where $t = 1, ..., T$; $i = 1, ..., N$; $h = 1, ..., H$. The question here is how to jointly estimate the parameters k_i^h and β_i^h within a 3-dimensional $(N \times T \times H)$ panel setup? Following Issler and Lima (2009), one can always decompose the series y_t into $\mathbb{E}_{t-h}(y_t)$ and an unforecastable component ζ_t^h, such that $\mathbb{E}_{t-h}(\zeta_t^h) = 0$, so that:

$$y_t = \mathbb{E}_{t-h}(y_t) + \zeta_t^h \tag{10}$$

Recall that $\eta_t^h = -\zeta_t^h$. Thus, it follows that:

$$f_{i,t}^{h} = k_i^h + \beta_i^h (y_t + \eta_t^h) + \varepsilon_{i,t}^{h} \tag{11}$$

Now, define $v_{i,t}^h \equiv \beta_i^h \eta_t^h + \varepsilon_{i,t}^h$ and assume that $\mathbb{E}(\varepsilon_{i,t}^h \mid \mathcal{F}_{t-h}) = 0$. This way, it follows that $\mathbb{E}\left(v_{i,t}^h \mid \mathcal{F}_{t-h}\right) = 0$ and that:

$$\mathbb{E}\left[\left(f_{i,t}^h - k_i^h - \beta_i^h y_t\right) \otimes z_{t-s}\right] = 0 \tag{12}$$

for all $i = 1, ..., N$ and $h = 1, ..., H$, where $z_{t-s} \in \mathcal{F}_{t-h}$ is a vector of instruments, such that $s \geqslant h$. The previous group of equations can be used as moment conditions (within a GMM setup) with $2NH$ parameters and (at least) $2NH$ moment conditions, provided that $dim(z_{t-s}) \geqslant 2$.

Nonetheless, as long as $N \to \infty$ the amount of parameters to estimate also diverges. Since our focus here is to construct a bias-correction device based on the aggregate estimates $\widehat{B^h}$ and $\widehat{\beta}^h$, one could reduce the problem dimensionality (e.g.

Driscoll and Kraay, 1998) by assuming (for instance) the following set of H moment conditions:

$$\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{i,t}^{h}\right)=0. \tag{13}$$

However, we follow here a different route, and estimate the $2H$ parameters of interest ($\widehat{B^h}$ and $\widehat{\beta}^h$) via GMM based on the two following setups:

Model 1 based on the individual (disaggregated) forecasts $f_{i,t}^h$ and the following moment conditions:

$$\mathbb{E}\left[\left(f_{i,t}^h - B^h - \beta^h y_t\right)\otimes z_{t-s}\right]=0 \tag{14}$$

for all $i = 1, ..., N$ and all $h = 1, ..., H$.

Model 2 based on the average forecast $\overline{f}_t^h \equiv \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$ and the following moment conditions:

$$\mathbb{E}\left[\left(\overline{f}_t^h - B^h - \beta^h y_t\right)\otimes z_{t-s}\right]=0 \tag{15}$$

for all $h = 1, ..., H$.

Notice that both models can be identified by using standard GMM estimates, as long as $dim(z_{t-s}) \geqslant 2$ and $s \geqslant h$. We next verify more formally the conditions necessary to construct an optimal forecast device and generate a proxy for the conditional expectation of y_t. Define $\theta_i^h = [k_i^h; \beta_i^h]'$ and consider the following assumptions:

Assumption A7 $(\theta_i^h, \eta_t^h, \varepsilon_{i,t}^h)$ are independent of each other for all i and t and a given h.

Assumption A8 θ_i^h is an identically distributed random vector in the cross-sectional dimension i, but not necessarily independent, i.e.,

$$\theta_i^h \sim \text{i.d.}(\theta^h, \Sigma_{\theta^h}), \tag{16}$$

where $\theta^h = [B^h; \beta^h]'$ and $\Sigma_{\theta^h} = \begin{bmatrix} \sigma_{k^h}^2 & \gamma_{k,\beta} \\ \gamma_{k,\beta} & \sigma_{\beta^h}^2 \end{bmatrix}$ such that $\sigma_{k^h}^2; \sigma_{\beta^h}^2; \left|\gamma_{k,\beta}\right| < \infty$. In the time-series dimension θ_i^h has no variation (i.e. vector of fixed parameters), but it is a random vector in the cross-section dimension.

Assumption A9 The aggregate shock η_t^h is a stationary and ergodic moving average (MA) process of order at most $h-1$, with zero mean and variance $\sigma^2_{\eta^h} < \infty$.

Assumption A10 Let $\boldsymbol{\varepsilon}_t^h = \left(\varepsilon_{1,t}^h, \varepsilon_{2,t}^h, \dots \varepsilon_{N,t}^h\right)'$ be a $N \times 1$ vector stacking the errors $\varepsilon_{i,t}^h$ associated with all possible forecasts. Assume that the vector process $\left\{\boldsymbol{\varepsilon}_t^h\right\}$ is covariance-stationary and ergodic for the first and second moments, uniformly on N, and that $\mathbb{E}_{t-h}\left(\varepsilon_{i,t}^h\right) = 0$ for all i and t and a given h. Furthermore, we follow Issler and Lima (2009) and also assume that

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \mathbb{E}\left(\varepsilon_{i,t}^{h\prime} \varepsilon_{j,t}^h\right)\right| = 0. \tag{17}$$

Because the forecasts are computed h-steps ahead, forecast errors $\varepsilon_{i,t}^h$ can be serially correlated. Assuming that $\varepsilon_{i,t}^h$ is weakly stationary is a way of controlling its time-series dependence. It does not rule out errors displaying conditional heteroskedasticity, since the latter can coexist with the assumption of weak stationarity.

Assumption A11 $\operatorname*{plim}_{N\to\infty} \left| \frac{1}{N} \sum_{i=1}^{N} \beta_i^h \right| < \infty$

Assumption A12 $\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}} \widehat{\theta}_i^h = \theta_i^h$; where $\widehat{\theta}_i^h$ are consistent GMM estimates, based on the moment conditions $\mathbb{E}\left[\left(f_{i,t}^h - k_i^h - \beta_i^h y_t\right) \otimes z_{t-s}\right] = 0$, where $z_{t-s} \in \mathcal{F}_{t-h}$ is a vector of instruments, such that $s \geqslant h$ and $dim(z_{t-s}) \geqslant 2$.

Proposition 8 *If A1-A4 and A7-A12 hold, then, the feasible Extended BCAF (Bias Corrected Average Forecast)* $\frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}}$ *based on GMM estimates* $\widehat{\theta}^h = [\widehat{B^h}; \widehat{\beta^h}]$ *obeys*

$$\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}} \right) = \mathbb{E}_{t-h}\left(y_t\right).$$

It is worth mentioning that our identification strategy is entirely based on GMM estimates. The validity of overidentifying restrictions can be checked by using standard TJ statistics. In our case, following Issler and Lima (2009), we rely on double asymptotics $(T, N \to \infty)$ and (at the same time) assume that the forecast horizon h remains bounded and, thus, does not diverge $(H < \infty)$, which is a reasonable hypothesis when dealing with actual survey forecasts (e.g. SPF, Michigan, Livingston).[7] Furthermore, the key identification restriction $\mathbb{E}\left(v_{i,t}^h \mid \mathcal{F}_{t-h}\right) = 0$ is essentially due to the orthogonal decomposition (10) plus a restriction on the error term $\varepsilon_{i,t}^h$, that is, $\mathbb{E}(\varepsilon_{i,t}^h \mid \mathcal{F}_{t-h}) = 0$.

On the other hand, the Extended BCAF is a generalization of the original BCAF proposed by Issler and Lima (2009), which considered $\beta^h = 1$. Notice that our setup also includes the Davies and Lahiri (1995) and Davies (2006) framework, reproduced below with our notation:

$$y_t - f_{i,t}^h = -\left(k_i^h + \eta_t^h + \varepsilon_{i,t}^h\right). \tag{18}$$

By imposing $\beta_i^h = 1$ for all $i = 1, ..., N$ and all $h = 1, ..., H$; in the extended BCAF setup, one can reproduce the three-dimensional error structure of the referred authors.

Finally, notice that the Extended BCAF is computed by $\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}}$, such that $\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}} \left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}}\right) = \mathbb{E}_{t-h}(y_t)$. In other words, despite the fact that individual forecasts $f_{i,t}^h$ might come (for instance) from asymmetric risk functions, the Extended BCAF is designed to preserve optimality under the MSE loss, as next discussed.

2.3 Optimality

In this section, we discuss the optimality of the original BCAF estimator and an implied testable restriction. First, notice that there exists a scalar Wold representation for y_t of the form:

$$y_t = \kappa_t + \sum_{j=0}^{\infty} \psi_j \xi_{t-j} = \kappa_t + \sum_{j=0}^{h-1} \psi_j \xi_{t-j} + \sum_{j=h}^{\infty} \psi_j \xi_{t-j}, \tag{19}$$

[7] Alternative identification strategies could be pursued (e.g. within-group method), which could potentially lead to more efficient estimates. For instance, one could explore the panel model with fixed effects proposed by Bai (2009), which allows for the joint presence of additive and interactive effects. Nonetheless, we leave this route as a suggestion for future research.

where κ_t is the deterministic term, $\psi_0 = 1$, $\sum_{j=1}^{\infty} \left|\psi_j\right| < \infty$ and ξ_{t-j} is white noise. By taking the conditional expectation h-periods before, it follows that:

$$\mathbb{E}_{t-h}(y_t) = \kappa_t + \sum_{j=0}^{h-1} \psi_j \mathbb{E}_{t-h}(\xi_{t-j}) + \sum_{j=h}^{\infty} \psi_j \mathbb{E}_{t-h}(\xi_{t-j}) = \kappa_t + \sum_{j=h}^{\infty} \psi_j \xi_{t-j}. \tag{20}$$

Therefore, we have that

$$y_t - \mathbb{E}_{t-h}(y_t) = \sum_{j=0}^{h-1} \psi_j \xi_{t-j}. \tag{21}$$

From equation (10) it also follows that $y_t - \mathbb{E}_{t-h}(y_t) = -\eta_t^h$ and, thus,

$$\eta_t^h = -\sum_{j=0}^{h-1} \psi_j \xi_{t-j}, \tag{22}$$

which is exactly assumption 3 of Issler and Lima (2009), in which the shock η_t^h is a MA process of order at most $h-1$, which is a testable restriction. Now, we discuss the optimality of the original BCAF estimator under a more general risk function. To do so, consider that $f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h$ and also that $y_t = \mathbb{E}_{t-h}(y_t) + \zeta_t^h = \mathbb{E}_{t-h}(y_t) - \eta_t^h$, and define $\beta^h \equiv \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \beta_i^h$, $\left|\beta_i^h\right| < \infty$ uniformly in i. Thus, it follows that

$$\begin{aligned}
f_{i,t}^h &= k_i^h + \beta_i^h y_t + \beta_i^h \eta_t^h + \varepsilon_{i,t}^h && (23)\\
\therefore \quad \frac{(f_{i,t}^h - k_i^h)}{\beta_i^h} &= y_t + \eta_t^h + \frac{\varepsilon_{i,t}^h}{\beta_i^h} && (24)\\
\therefore \quad \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{(f_{i,t}^h - k_i^h)}{\beta_i^h} \right) &= y_t + \eta_t^h && (25)\\
\therefore \quad \mathbb{E}[\underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{(f_{i,t}^h - k_i^h)}{\beta_i^h} \right) &- y_t]^2 = \sigma_{\eta^h}^2. && (26)
\end{aligned}$$

Notice that under the original BCAF of Issler and Lima (2009) and ignoring the existence of the slope coefficient β_i^h it follows that

$$\frac{1}{N}\sum_{i=1}^{N}(f_{i,t}^h - k_i^h) \xrightarrow{p} \beta^h y_t + \beta^h \eta_t^h \tag{27}$$

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}(f_{i,t}^h - k_i^h - y_t) &\xrightarrow{p} (\beta^h - 1)y_t + \beta^h \eta_t^h && (28) \\ = \quad & (\beta^h - 1)y_t + (\beta^h - 1)\eta_t^h + \eta_t^h && (29) \\ = \quad & (\beta^h - 1)\mathbb{E}_{t-h}(y_t) + \eta_t^h && (30) \end{aligned}$$

This way, it follows that:

$$\mathbb{E}[\operatorname*{plim}_{N\to\infty}(\frac{1}{N}\sum_{i=1}^{N}(f_{i,t}^h - k_i^h)) - y_t]^2 \tag{31}$$

$$= \quad (\beta^h - 1)^2(\mathbb{E}_{t-h}(y_t))^2 + \sigma_{\eta^h}^2 > \sigma_{\eta^h}^2. \tag{32}$$

Therefore, if the forecast bias is not only additive, i.e. $f_{i,t}^h = \beta_i^h \mathbb{E}_{t-h}(y_t) + k_i^h + \varepsilon_{i,t}^h$, where $\beta_i^h \sim \left(\beta^h, \sigma_{\beta^h}^2\right)$, such that $\beta^h \neq 1$, the BCAF is no longer optimal. Optimality can be restored if the BCAF is slightly modified to be $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{f_{i,t} - \widehat{k_i^h}}{\widehat{\beta_i^h}}\right)$, where $\widehat{k_i^h}$ and $\widehat{\beta_i^h}$ are consistent estimators of k_i^h and β_i^h, respectively.

2.4 Aggregate Forecasts

In this section, we remind the reader that OLS estimates from a standard (intercept-slope) bias correction model based on aggregated forecasts (Elliott et al., 2006) might not be necessarily unbiased. To investigate this issue more carefully, first consider the econometric model under Case 3 and Proposition 8. Then, take the cross-section average on both sides of $f_{i,t}^h = k_i^h + \beta_i^h(y_t + \eta_t^h) + \varepsilon_{i,t}^h$, such that:

$$\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h = \frac{1}{N}\sum_{i=1}^{N}\left(k_i^h + \beta_i^h(y_t + \eta_t^h) + \varepsilon_{i,t}^h\right) \tag{33}$$

$$\therefore \quad y_t = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(f_{i,t}^h - k_i^h - \beta_i^h\eta_t^h - \varepsilon_{i,t}^h\right)}{\beta_i^h} \tag{34}$$

One can rewrite the last equation in the following way:

$$y_t = c_0^h + c_1^h \overline{f}_t^h + \epsilon_t^h \tag{35}$$

where $c_0^h = \frac{1}{N}\sum_{i=1}^{N}\frac{-k_i^h}{\beta_i^h}$; $c_1^h = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\beta_i^h}$; $\overline{f}_t^h \equiv \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$; $\epsilon_t^h = \frac{1}{N}\sum_{i=1}^{N}\left(-\eta_t^h - \frac{\varepsilon_{i,t}^h}{\beta_i^h}\right)$. Based on the time series $\{y_t; \overline{f}_t^h\}_{t=1}^T$, one could employ OLS to estimate parameters $(c_0^h; c_1^h)'$ or, alternatively, use Nonlinear Least Squares (NLS) to directly estimate parameters $(B^h; \beta^h)'$ from the regression:

$$y_t = -\frac{B^h}{\beta^h} + \frac{1}{\beta^h}\overline{f}_t^h + \epsilon_t^h. \tag{36}$$

Notice that it is a linear regression with nonlinear restriction on the coefficients. Nonetheless, as well known, if $\mathbb{E}(\overline{f}_t^{h\,\prime}\epsilon_t^h) \neq 0$, then, OLS estimator might not be unbiased. To check it more carefully, consider the following double asymptotics $(T, N \to \infty)_{\text{seq}}$ and assume that h remains bounded, as it follows:

$$\begin{aligned} \lim_{(T,N\to\infty)_{\text{seq}}} \mathbb{E}(\epsilon_t^h) &= \lim_{(T,N\to\infty)_{\text{seq}}} -\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\eta_t^h + \frac{\varepsilon_{i,t}^h}{\beta_i^h}\right) & (37)\\ &= \lim_{(T,N\to\infty)_{\text{seq}}} -\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}(\eta_t^h) + \frac{\mathbb{E}(\varepsilon_{i,t}^h)}{\mathbb{E}(\beta_i^h)}\right) = 0, & (38)\end{aligned}$$

where the last two equalities are due to $\mathbb{E}(\eta_t^h) = 0$ for all t and h; $\mathbb{E}(\varepsilon_{i,t}^h) = 0$ for all i and t; $(\beta_i^h, \varepsilon_{i,t}^h)$ are independent of each other for all i and t and a given h; and $\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}} \frac{1}{N}\sum_{i=1}^{N}\beta_i^h = \beta^h$; $|\beta^h| < \infty$. We also assume that $\beta^h \neq 0$ (to exclude the non-interesting case in which the aggregate forecast would be just a random error around B^h).

In addition, notice that:

$$\begin{aligned} \mathbb{E}(\epsilon_t^{h\prime}\epsilon_t^h) &= \mathbb{E}\left(\left(-\frac{1}{N}\sum_{i=1}^{N}\left(\eta_t^h + \frac{\varepsilon_{i,t}^h}{\beta_i^h}\right)\right)'\left(-\frac{1}{N}\sum_{i=1}^{N}\left(\eta_t^h + \frac{\varepsilon_{i,t}^h}{\beta_i^h}\right)\right)\right) & (39)\\ \lim_{(T,N\to\infty)_{\text{seq}}} \mathbb{E}(\epsilon_t^{h\prime}\epsilon_t^h) &= \lim_{(T,N\to\infty)_{\text{seq}}} \frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}(\eta_t^{h\prime}\eta_t^h) + \frac{2\mathbb{E}(\eta_t^{h\prime}\varepsilon_{i,t}^h)}{\mathbb{E}(\beta_i^h)} + \frac{\mathbb{E}(\varepsilon_{i,t}^{h\prime}\varepsilon_{i,t}^h)}{\mathbb{E}(\beta_i^{h\prime}\beta_i^h)}\right) & (40)\\ &= \sigma_{\eta^h}^2 > 0 & (41)\end{aligned}$$

since $\mathbb{E}(\eta_t^{h\prime}\eta_t^h) = \sigma^2_{\eta^h}$; $(\eta_t^h, \varepsilon_{i,t}^h, \beta_i^h)$ are independent of each other for all i, t and a given h; and $\lim_{(T,N\to\infty)_{\text{seq}}} \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(\varepsilon_{i,t}^{h\prime}\varepsilon_{i,t}^h) = 0$ by following the proof of Lemma 1 of Issler and Lima (2009). Moreover, provided that $\lim_{(T,N\to\infty)_{\text{seq}}} \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(\beta_i^h) = \beta^h \neq 0$, by assumption, and since $\lim_{(T,N\to\infty)_{\text{seq}}} \frac{1}{N}\sum_{i=1}^{N}VAR(\beta_i^h) = \sigma^2_{\beta^h} > 0$, it also follows that $\lim_{(T,N\to\infty)_{\text{seq}}} \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(\beta_i^{h\prime}\beta_i^h) = \left(\sigma^2_{\beta^h} + \left(\beta^h\right)^2\right) \neq 0$. Now, let's investigate the term $\mathbb{E}(\overline{f}_t^{h\,\prime}\epsilon_t^h)$.

$$\begin{aligned}
\mathbb{E}(\overline{f}_t^{h\,\prime}\epsilon_t^h) &= \mathbb{E}\left(\left(\frac{1}{N}\sum_{i=1}^{N}(k_i^h + \beta_i^h y_t) + (\beta_i^h\eta_t^h + \varepsilon_{i,t}^h)\right)'\epsilon_t^h\right) && (42)\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(k_i^h\mathbb{E}\left(\epsilon_t^h\right) + \beta_i^h\mathbb{E}\left(y_t'\epsilon_t^h\right) - \beta_i^h\mathbb{E}\left(\epsilon_t^{h\prime}\epsilon_t^h\right)\right) && (43)
\end{aligned}$$

$$\begin{aligned}
\therefore \lim_{(T,N\to\infty)_{\text{seq}}} \mathbb{E}(\overline{f}_t^{h\,\prime}\epsilon_t^h) &= \operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}}\left(\frac{1}{N}\sum_{i=1}^{N}\beta_i^h\mathbb{E}\left(y_t'\epsilon_t^h\right)\right) - \beta^h\sigma^2_{\eta^h} && (44)\\
= \operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}}&\left(\frac{1}{N}\sum_{i=1}^{N} - \beta_i^h\mathbb{E}\left(y_t'\eta_t^h\right) - \mathbb{E}\left(y_t'\varepsilon_{i,t}^h\right)\right) - \beta^h\sigma^2_{\eta^h} && (45)\\
= -\lim_{(T,N\to\infty)_{\text{seq}}}&\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(y_t'\varepsilon_{i,t}^h) && (46)
\end{aligned}$$

where $\mathbb{E}(y_t'\eta_t^h) = -\sigma^2_{\eta^h}$ comes from the orthogonal decomposition: $y_t = \mathbb{E}_{t-h}(y_t) - \eta_t^h$. Since $\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h = 0$ (with variance equal to zero, in the limit) it follows that $\lim_{(T,N\to\infty)_{\text{seq}}}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(y_t'\varepsilon_{i,t}^h) = 0$ (recall that y_t is assumed to be a weakly stationary and ergodic process).

This way, the OLS condition $\mathbb{E}(\overline{f}_t^{h\,\prime}\epsilon_t^h) = 0$ can be asymptotically achieved (although requiring double asymptotics). Nonetheless, in cases where only the time dimension diverges (i.e. $T \to \infty$ with small N) biased OLS coefficients can be generated. An alternative approach to overcome this potential (finite sample) bias is to employ a standard instrumental variable (IV) setup by assuming, for instance, that $\mathbb{E}(z_{t-s}'\epsilon_t^h) = 0$, for $s \geqslant h$, based on a set of valid instruments z_t. Notice that this approach can be nested into the GMM setup discussed in previous sections.[8] In

[8] In fact, recall that the GMM setup allows for moment conditions other than the considered in

the empirical exercise, we employ aggregate forecasts to investigate bias-correction based on the (GMM) moment conditions (15) and the Nonlinear Least Squares (NLS) regression (36).

3 Empirical Application

3.1 Data

In this section, we employ the BCAF of Issler and Lima (2009) to a varying forecast horizon, based on Brazilian inflation data, and compare it to our proposed extended setup. In particular, we focus our analysis on the behavior of forecasts of the monthly inflation rate in Brazil, as measured by IPCA, which is a consumer price index (CPI) used to compute the official inflation target. Our goal is to estimate the "forecast error term-structure" of the inflation rate based on market forecasts obtained from the Focus survey.

The Focus survey[9] is a unique panel database of forecasts, organized by the Central Bank of Brazil, which collects daily information on almost 100 institutions, including commercial banks, asset-management firms, and non-financial institutions, which are followed throughout time with a reasonable turnover. Forecasts are computed at different frequencies, for a large array of macroeconomic time series included in the survey, as well as at different forecast horizons, which potentially can serve to approximate a large N, T environment for techniques designed to deal with unbalanced panels – which is not the case studied here. Besides the large size of N and T, the Focus survey also has the following desirable features: the anonymity of forecasters is preserved (i.e. there is no reputation concerns and design of competition is likely to encourage truthful reporting), although the list of the top-five forecasters for a given economic variable is periodically released by the Central Bank of Brazil (see Carvalho and Minella, 2012).

this paper, for instance, focused on higher moments of the forecast error, or even mixing coefficients from distinct forecast horizons (e.g. moments based on forecast revisions), which might useful to smooth the estimated coefficients along the horizons. In addition, the aggregate approach can be less efficient since it employs much less data in comparison to the disaggregated counterpart.

[9] The collection and manipulation of data from the Focus survey is conducted exclusively by the staff of the Central Bank of Brazil.

In order to obtain the largest possible balanced panel ($N \times T \times H$), we used a time-series sample period covering daily forecasts collected from 02 January 2006 to 07 February 2014. Every day considered within the sample, market agents $i = 1, ..., N$ inform their expectations regarding the inflation rate for the next 14 months. For instance, market agent i informs on 05 January 2006 his (or her) forecast for inflation rate in December 2005 (not yet released), as well as for January 2006 and the following 12 months. Next day, the same agent updates the forecasts for the same inflation rates in December 2005, January 2006,..., January 2007.

This way, our sample covers forecasts for the CPI inflation rates from December 2005 to January 2014 ($t = 1, ..., T = 98$ months), which represents a period of stable inflation in Brazil (see Figure 1), and the forecast horizons range from $h = 1, ..., H = 400$ days. This way, the original database represents forecasts for "fixed-events" (see Bakhshi et al., 2005) and varying forecast horizons. These original forecasts are re-organized to form time series of fixed-forecast-horizons and time-varying-events. As a result, the dataset forms an unbalanced panel ($N \times T \times H$) containing an amount of 2,732,827 observations. The final dataset used in this paper contains 1,486,559 observations, since we only consider the forecasts from agents that participate on the survey in a regular basis (i.e., forming a balanced panel). Figure 3 shows the number of effective agents N for each horizon h.

We consider two consecutive time sub-periods, where the first sub-period ($t = 1, ..., T_1$) is labeled "training sample", where realizations of y_t are usually confronted with forecasts provided by the survey of forecasts, and bias-correction terms are estimated. The second sub-period is where genuine out-of-sample forecast is entertained, comprising the last P observations ($t = T_1 + 1, ..., T$) to compute $MSE_h = \frac{1}{P} \sum_{t=T_1+1}^{T} \left(y_t - \widehat{f}_t^h \right)^2$, where $\widehat{f}_t^h$ is either the BCAF or the Extended BCAF forecast. For the average forecast, there is no parameter to be estimated using training-sample observations. Out-of-sample forecasts are computed according to $f_t^{average,h} = \frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h$, $t = T_1 + 1, ..., T$, and its MSE is similarly computed. In this sense, we chose the first $R = 60$ time observations to estimate the model parameters leaving $P = 38$ time-series observations for (pseudo) out-of-sample forecast evaluation.

We estimate both the original BCAF of Issler and Lima as well as the proposed Extended BCAF. Regarding the original BCAF, we estimate the market bias $\widehat{B^h}$ for each horizon h separately. In respect to the Extended BCAF, we use a set of instruments containing lagged figures of inflation π_t and interest i_t rates (IPCA and Selic, respectively), nominal exchange rate (R$/US$) and industrial production. Interest rate, nominal exchange rate and industrial production index are all first-differenced and log-transformed. The reported results are based on the following set of instruments: $z_{t-s} = [1; \pi_{t-s}; \pi_{t-s-2}; \pi_{t-s-5}; i_{t-s-5}]'$ with $s = 14$ months.

Notice that our dataset (in fact) represents an unbalanced 3-dimensional setup with mixed frequencies (t is measured in a monthly frequency, whereas h is provided in a daily basis). The ideal situation would be to estimate the Extended BCAF by jointly using the 3 dimensions altogether. However, in practice, this estimation was not feasible (i.e. singularities in the numerical optimization arise), probably due to the fact that forecasts of two consecutive horizons might be highly correlated.[10] Joint (3D) estimations with selected horizons (e.g. $h = 30, 60, 90, ..., 180$ days) also resulted in convergence problems. The solution adopted was to estimate the model for each horizon h separately (as done in the original BCAF). In this sense, we considered Models 1 and 2 (described in Case 3 of Section 2.2) by using individual as well as the average forecasts.

The empirical exercise is all conducted by using the R software (version 3.0.1) and the package "gmm". The covariance matrix of coefficients is estimated in the original BCAF by using the "random fields" approach (Conley, 1999) as in Issler and Lima (2009), whereas in the Extended BCAF it is estimated by using the "two step" approach of Hansen (1982), although the "iterative" procedure of Hansen et al. (1996) leads to similar results.

[10] On the other hand, some individuals in the survey may infrequently update their forecasts; which could also generate singularities in the optimization problem. A possible solution, in this case, would be to estimate the model with monthly data in the forecast horizon dimension.

3.2 Results

The results of our empirical exercise are presented in Tables and Figures in Appendix. The results in Table 1 show that the average bias is negative for the whole set of considered horizons (suggesting that, in general, survey participants underestimate inflation) and statistically significant after a 6-month horizon. The estimated bias for a 7-month horizon (-0.0456) is approximately -0.55% in a yearly basis, which is a sizable bias for an average inflation rate around 5.6% in twelve months (as of January 2014).

In Table 2, pseudo out-of-sample forecast comparisons between the average forecast and the original bias-corrected average forecast show that the former has a MSE 16% bigger than that of the latter when considering a twelve-month forecast horizon, which shows that the forecasting gains from bias correction might be non-trivial. It is worth mentioning that these results are in line with the previous findings of Issler and Lima (2009), which report a p-value of 0.063 for a null-bias test with a 6-month horizon, despite their different sample (from November 2002 to March 2006) in comparison to the one used in our empirical exercise (from December 2005 to January 2014).

Figure 4 shows the so-called forecast error "term-structure" estimated via the original BCAF of Issler-Lima. Notice that, as expected, the market bias collapses to zero (i.e., the BCAF converges to the simple average forecast) as long as the forecast horizon approaches to zero.

In Appendix A.4, the results for the Extended BCAF are presented. Table 3 presents the comparison between models 1 and 2, leading to similar point estimates, although model 1 (with individual forecasts), as expected, presents lower confidence intervals for estimated coefficients. In both cases, the Wald test of "no bias" is strongly rejected in all considered horizons, in sharp contrast to the respective results from the original BCAF, which pointed out to the existence of forecast bias only after a 6-month horizon. To save space, the remaining Figures and Tables only show the results for model 2 (quite similar to the results from model 1).

Figures 9 and 10 present the estimated coefficients along the daily forecast horizons h, revealing the evolution of forecast-bias along the "term-structure" in the short-run (up to 90 days), which seems to remain relatively constant in the medium to long-run.

On the other hand, notice that the possibility of a more flexible model, allowing for the slope coefficient (associated to the conditional expectation) to be different from unity, indeed leads to sizable out-of-sample forecast improvements. Figures 11 and 12 show that the Extended BCAF generates MSE ratios (in respect to the average forecasts) with gains greater than those produced by the original BCAF.

Indeed, the equal predictive accuracy test of Clark and West (2007) for nested models (Figure 13) suggests that both bias-correction devices can statistically reduce (at a 10% significance level) out-of-sample MSE for horizons above 9 months (and marginally reduce the MSE for h ranging between 1 and 3 months). The Clark-West test also indicates that the extended BCAF can statistically improve out-of-sample predictability over the original BCAF for longer horizons ($h > 10$ months).

Regarding model specification, Table 4 shows the p-values of a testable restriction that the aggregate shock η_t^h, estimated within the extended setup, follows a moving average process of order (at most) $h-1$. Overall, the results indicate that the aggregate shock seems not to violate the optimal restriction of MA($h-1$).

Finally, in Appendix A.5 we present a comparison between Nonlinear Least Squares (NLS) and GMM estimates, both based on aggregate forecasts. Notice that distinct estimation techniques lead to quite different in-sample results.[11] Moreover, the out-of-sample performance of the estimated bias-correction devices suggests that GMM generates lower MSEs in almost all horizons in comparison to NLS. In addition, according to the Clark-West (2007) test, the NLS-based model generates MSEs no better than the MSEs from the standard average forecast (excepting a few marginal results for horizons between 1 and 3 months).

4 Conclusion

In this paper, we use an econometric approach to forecast stationary and ergodic series y_t within a 3-dimensional panel-data framework, where the number of forecasts and the number of time periods increase without bounds. Our method is linked

[11] A Hausman specification test could be employed to further investigate this difference.

to the previous literature on two-way (and three-way) error decomposition (e.g. Wallace and Hussain (1969); Fuller and Battese (1974); Davies and Lahiri (1995)) as well as on forecast combination and bias-correction devices (e.g. Granger and Ramanathan (1984); Issler and Lima (2009)). As shown here, standard tools from panel-data asymptotic theory are used to devise an optimal forecast combination that delivers $\mathbb{E}_{t-h}(y_t)$ under more general loss functions than usual MSE.

The novelty in this paper is to propose a bias-correction device for a whole set of forecast horizons, in order to reveal the forecast error "term-structure", that is, the size of aggregate bias in respect to the forecast horizon. More importantly, a microfounded framework is presented to justify the existence of a forecast bias, based on the optimization problem of individual forecasters. Among the many potential sources of forecast bias (e.g. model misspecification, distinct information sets among forecasters due to private information, asymmetric loss functions and its relationship with the DGP), we focus here on the later source and present a detailed discussion about asymmetric loss functions; which might lead the standard bias-correction approach (based only on intercept correction) to become sub-optimal in the presence of such asymmetries from individual forecasters.

In such a context, a GMM setup is suggested to generate a feasible optimal (combined) forecast, constructed to deliver the conditional expectation of the target variable. An empirical exercise using a dataset of Brazilian inflation expectations illustrates our methodology, suggesting that the proposed model is able to produce out-of-sample forecasts with superior performance (in the MSE sense), for several forecast horizons, in comparison to the average forecast and to the additive-only bias-correction approach of Issler and Lima (2009).

Possible extensions of this paper include: (i) investigate the dynamics, and related statistical properties, of the forecast revision process $R_{i,t}^h \equiv f_{i,t}^h - f_{i,t}^{h+1}$; (ii) pursue other identification strategies and conduct a joint 3-dimensional estimation; and (iii) tackle the unbalanced panel issues (e.g. missing values, ragged edge data) by using, for instance, the Expectation Maximization (EM) algorithm (Stock and Watson, 2002a,b).

References

[1] Ang, A., Bekaert, G., Wei, M., 2007. Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better? *Journal of Monetary Economics* 54, 1163-1212.

[2] Bai, J., 2009. Panel Data Models with Interactive Fixed Effects. *Econometrica* 77(4), 1229-1279.

[3] Bakhshi, H., Kapetanios, G., Yates, T., 2005. Rational expectations and fixed-event forecasts: An application to UK inflation. *Empirical Economics* 30, 539-553.

[4] Batchelor, R., 2007. Bias in macroeconomic forecasts. *International Journal of Forecasting* 23, 189-203.

[5] Bates, J.M., Granger, C.W.J., 1969. The Combination of Forecasts. *Operations Research Quarterly* 20, 309-325.

[6] Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics* 167 (2), 413-425.

[7] Capistrán, C., Timmermann, A., 2009. Forecast Combination with Entry and Exit of Experts. *Journal of Business and Economic Statistics* 27, 428-40.

[8] Carvalho, F.A., Minella, A., 2012. Survey Forecasts in Brazil: A Prismatic Assessment of Epidemiology, Performance, and Determinants. *Journal of International Money and Finance* 31(6), 1371-1391.

[9] Clark, T.E., West, K.D., 2007. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138, 291-311.

[10] Coibion, O., Gorodnichenko, Y., 2012. What can survey forecasts tell us about information rigidities? *Journal of Political Economy* 120, 116-159.

[11] Conley, T.G., 1999. GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics* 92, 1-45.

[12] Davies, A., 2006. A Framework for Decomposing Shocks and Measuring Volatilities Derived from Multi-Dimensional Panel Data of Survey Forecasts. *International Journal of Forecasting* 22(2), 373-393.

[13] Davies, A., Lahiri, K., 1995. A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics* 68, 205-227.

[14] Driscoll, J., Kraay, A., 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80, 549-560.

[15] Elliott, G., Granger, C.W.J., Timmermann, A., 2006. Editors, *Handbook of Economic Forecasting*, Amsterdam: North-Holland.

[16] Elliott, G., Komunjer, I., Timmermann, A. 2008. Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss? *Journal of the European Economic Association* 6(1), 122-157.

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

[18] Faust, J., Wright, J.H., 2012. Forecasting Inflation. Mimeo. Available at http://www.econ2.jhu.edu/people/wright/inf_forecast.pdf

[19] Fuller, W.A., Battese, G.E., 1974. Estimation of linear models with crossed-error structure. *Journal of Econometrics* 2, 67-78.

[20] Gaglianone, W.P., Lima, L.R., 2012. Constructing Density Forecasts from Quantile Regressions. *Journal of Money, Credit and Banking* 44(8), 1589-1607.

[21] Galvao, A.F., Wang, L., 2013. Efficient minimum distance estimator for quantile regression fixed effects panel data. Working Paper.

[22] Granger, C.W.J., 1969. Prediction with a generalized cost of error function. *Operational Research Quarterly* 20 (2), 199-207.

[23] Granger, C.W.J., Newbold, P., 1986. Forecasting time series, 2nd Edition. Academic Press.

[24] Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasting. *Journal of Forecasting* 3, 197–204.

[25] Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-1054.

[26] Hansen, L.P., Heaton, J., Yaron, A., 1996. Finit-Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics* 14, 262-280.

[27] Hendry, D.F., Clements, M.P., 2002. Pooling of forecasts. *Econometrics Journal* 5, 1-26.

[28] Issler, J.V., Lima, L.R., 2009. A Panel Data Approach to Economic Forecasting: The Bias-corrected Average Forecast. *Journal of Econometrics* 152(2), 153 – 164.

[29] Julio, J.M., 2007. The fan chart: The technical details of the new implementation. Borradores de Economía n. 468. Banco de la República. Available at http://www.banrep.gov.co/docum/ftp/borra468.pdf

[30] Koenker, R., 2005. Quantile Regression. Cambridge University Press.

[31] Koenker, R., Bassett, G., 1978. Regression Quantiles. *Econometrica* 46, 33-50.

[32] Lahiri, K., Peng, H., Sheng, X., 2013. Measuring Uncertainty of a Combined Forecast and a New Test for Forecaster Homogeneity. Mimeo.

[33] Laster, D., Bennett, P., Geoum, I.S., 1999. Rational Bias In Macroeconomic Forecasts. *The Quarterly Journal of Economics* 114, 293-318.

[34] Mankiw, N.G., Reis, R., 2010. Imperfect Information and Aggregate Supply. NBER Working Paper No. 15773.

[35] Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handbook of Econometrics. In: R.F. Engle & D. McFadden (ed.), vol.IV, Ch.36. Elsevier Science, 2111-2245.

[36] Ottaviani, M., Sorensen, P.N., 2006. The strategy of professional forecasting. *Journal of Financial Economics* 81, 441-466.

[37] Palm, F.C., Zellner, A., 1992. To combine or not to combine? Issues of combining forecasts. *Journal of Forecasting* 11(8), 687-701.

[38] Patton, A.J., Timmermann, A., 2007. Testing Forecast Optimality under Unknown Loss. *Journal of the American Statistical Association* 102(480), 1172-1184.

[39] Phillips, P.C.B., Moon, H.R., 1999. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* 67, 1057-1111.

[40] Stock, J., Watson, M., 1999. Forecasting Inflation. *Journal of Monetary Economics* 44, 293-335.

[41] _ _ _ _ _, 2002a. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics* 20, 147-162.

[42] _ _ _ _ _, 2002b. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97(460), 1167-1179.

[43] _ _ _ _ _, 2006. Forecasting with Many Predictors. In: Elliott, G., C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, (Amsterdam: North-Holland), 515-554.

[44] Timmermann, A., 2006. Forecast Combinations. In: Elliott, G., C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, (Amsterdam: North-Holland) 135-196.

[45] Wallace, H.D., Hussain, A., 1969. The use of error components model in combining cross-section and time-series data. *Econometrica* 37, 55-72.

A Appendix

A.1 Proofs of Propositions

Proof of Proposition 1. Define the objective function $\Phi(\widetilde{y}) \equiv \mathbb{E}\left(L(y-\widetilde{y}) \mid \mathcal{F}_{t-h}\right)$ $= \int_{-\infty}^{\infty} L(y-\widetilde{y})f(y)dy = \int_{-\infty}^{\widetilde{y}} \beta_1 g(y-\widetilde{y})f(y)dy + \int_{\widetilde{y}}^{\infty} \beta_2 g(y-\widetilde{y})f(y)dy$, where the last equality is due to A3. Differentiating $\Phi(\widetilde{y})$ with respect to $\widetilde{y}$ and equating it to zero gives us the optimal forecast $\widetilde{y}^*$ and the respective value of the minimum expected loss $\Phi(\widetilde{y}^*)$. Thus, it follows that $\partial\Phi(\widetilde{y})/\partial\widetilde{y} = 0 \therefore \beta_1 \int_{-\infty}^{\widetilde{y}^*} g'(y-\widetilde{y}^*)f(y)dy = -\beta_2 \int_{\widetilde{y}^*}^{\infty} g'(y-\widetilde{y}^*)f(y)dy$. Recall that (by A3) $\{\beta_1;\beta_2\} > 0$, and denote $\gamma_1 \equiv \int_{-\infty}^{\widetilde{y}^*} g'(y-\widetilde{y}^*)f(y)dy \leq 0$, since $f(y) \geq 0$ for all $y \in \mathbb{R}$ and, based on assumptions A2-A3, $g'(e) \leq 0$ for all $e \leq 0$, which is equivalent to $y \in (-\infty;\widetilde{y}^*]$. In addition, denote $\gamma_2 \equiv \int_{\widetilde{y}^*}^{\infty} g'(y-\widetilde{y}^*)f(y)dy \geq 0$, since $f(y) \geq 0$ for all $y \in \mathbb{R}$ and, also based on assumptions A2-A3, $g'(e) \geq 0$ for all $e \geq 0$, which is equivalent to $y \in [\widetilde{y}^*;\infty)$. Thus $\beta_1\gamma_1 = -\beta_2\gamma_2$ or $\beta_1(-\gamma_1) = \beta_2\gamma_2 \geq 0$. (ii) If $\beta_1 > \beta_2$, then, $(-\gamma_1) < \gamma_2$. This way, $\int_{-\infty}^{\widetilde{y}^*} -g'(y-\widetilde{y}^*)f(y)dy < \int_{\widetilde{y}^*}^{\infty} g'(y-\widetilde{y}^*)f(y)dy$. Since (by A3) $g(e)$ is a symmetric function about $e=0$, that is $g(e) = g(-e)$, and its first derivative $g'(e)$ exists almost everywhere, then, it follows that $g'(e)$ must be an antisymmetric function about $e=0$, or $y=\widetilde{y}^*$, so that $g'(e) = -g'(-e)$. One can rewrite the previous inequality as $\int_{-\infty}^{\widetilde{y}^*} g_1(y-\widetilde{y}^*)f(y)dy < \int_{\widetilde{y}^*}^{\infty} g_1(y-\widetilde{y}^*)f(y)dy$, where $g_1(e) \equiv \begin{cases} -g'(e) & ; \; e<0 \\ g'(e) & ; \; e>0 \end{cases}$ is a symmetric and non-negative function around $e=0$, or $y=\widetilde{y}^*$, which implies (by using the symmetry of g_1 and a mean value theorem for integration) that $\int_{-\infty}^{\widetilde{y}^*} f(y)dy < \int_{\widetilde{y}^*}^{\infty} f(y)dy \therefore \int_{-\infty}^{\widetilde{y}^*} f(y)dy < 1 - \int_{-\infty}^{\widetilde{y}^*} f(y)dy$ since (by definition of a pdf) we have that $\int_{-\infty}^{\infty} f(y)dy = 1$. Thus, $\int_{-\infty}^{\widetilde{y}^*} f(y)dy < 0.5$ $\therefore \int_{-\infty}^{\widetilde{y}^*} f(y)dy = F_{t,t-h}(\widetilde{y}^*) \equiv \tau_i < 0.5 \therefore \widetilde{f}_{i,t}^h < Med_{t-h}(y_t)$.

(iii) If $\beta_1 < \beta_2$, then, based on a similar argument, it follows that $\widetilde{f}_{i,t}^h > Med_{t-h}(y_t)$. (i) Now, it trivially follows that if $\beta_1 \neq \beta_2$ then $F_{t,t-h}(\widetilde{y}^*) \neq 0.5$. (iv) Since $\beta_1^j/\beta_2^j < 1 \therefore \widetilde{f}_{j,t}^h > Med_{t-h}(y_t)$. Since $\beta_1(-\gamma_1) = \beta_2\gamma_2 \geq 0$, and if

$\beta_1^i/\beta_2^i < \beta_1^j/\beta_2^j$, then, $\gamma_2^i/\left(-\gamma_1^i\right) < \gamma_2^j/\left(-\gamma_1^j\right) \therefore -\gamma_1^j\gamma_2^i < -\gamma_1^i\gamma_2^j \therefore \int\limits_{-\infty}^{\widetilde{y_j}^*} - g'(y - \widetilde{y_j}^*)f(y)dy \int\limits_{\widetilde{y_i}^*}^{\infty} g'(y - \widetilde{y_i}^*)f(y)dy < \int\limits_{-\infty}^{\widetilde{y_i}^*} - g'(y - \widetilde{y_i}^*)f(y)dy \int\limits_{\widetilde{y_j}^*}^{\infty} g'(y - \widetilde{y_j}^*)f(y)dy$. Thus, $\int\limits_{-\infty}^{\widetilde{y_j}^*} g_1(y - \widetilde{y_j}^*)f(y)dy \int\limits_{\widetilde{y_i}^*}^{\infty} g'(y - \widetilde{y_i}^*)f(y)dy < \int\limits_{-\infty}^{\widetilde{y_i}^*} -g'(y - \widetilde{y_i}^*)f(y)dy \int\limits_{\widetilde{y_j}^*}^{\infty} g_1(y - \widetilde{y_j}^*)f(y)dy$

$$\therefore \int\limits_{-\infty}^{\widetilde{y_j}^*} g_1(y - \widetilde{y_j}^*)f(y)dy \int\limits_{\widetilde{y_i}^*}^{\infty} g_1(y - \widetilde{y_i}^*)f(y)dy < \int\limits_{-\infty}^{\widetilde{y_i}^*} g_1(y - \widetilde{y_i}^*)f(y)dy \int\limits_{\widetilde{y_j}^*}^{\infty} g_1(y - \widetilde{y_j}^*)f(y)dy.$$

Since $g_1(e_j)$ is a symmetric function about $e_j = 0$, or $y = \widetilde{y_j}^*$, following proof in (ii), it follows that $\int\limits_{-\infty}^{\widetilde{y_j}^*} f(y)dy \int\limits_{\widetilde{y_i}^*}^{\infty} g_1(y - \widetilde{y_i}^*)f(y)dy < \int\limits_{-\infty}^{\widetilde{y_i}^*} g_1(y - \widetilde{y_i}^*)f(y)dy \int\limits_{\widetilde{y_j}^*}^{\infty} f(y)dy$. Now , since $g_1(e_i)$ is a symmetric function about $e_i = 0$, or $y = \widetilde{y_i}^*$ it also follows that $\int\limits_{-\infty}^{\widetilde{y_j}^*} f(y)dy \int\limits_{\widetilde{y_i}^*}^{\infty} f(y)dy < \int\limits_{-\infty}^{\widetilde{y_i}^*} f(y)dy \int\limits_{\widetilde{y_j}^*}^{\infty} f(y)dy \therefore \int\limits_{-\infty}^{\widetilde{y_j}^*} f(y)dy \left(1 - \int\limits_{-\infty}^{\widetilde{y_i}^*} f(y)dy\right) < \int\limits_{-\infty}^{\widetilde{y_i}^*} f(y)dy \left(1 - \int\limits_{-\infty}^{\widetilde{y_j}^*} f(y)dy\right) \therefore \int\limits_{-\infty}^{\widetilde{y_j}^*} f(y)dy < \int\limits_{-\infty}^{\widetilde{y_i}^*} f(y)dy \therefore F_{t,t-h}(\widetilde{y_j}^*) < F_{t,t-h}(\widetilde{y_i}^*) \therefore \widetilde{f}_{j,t}^h < \widetilde{f}_{i,t}^h$ and from (iii) $\widetilde{f}_{j,t}^h > Med_{t-h}(y_t) \therefore \widetilde{f}_{i,t}^h > \widetilde{f}_{j,t}^h > Med_{t-h}(y_t)$. ■

Proof of Corollary 2. Based on Proposition 1, we have the nice result that if $\beta_1 > \beta_2$ then $\tau_i < 0.5$, but if $\beta_1 < \beta_2$ then $\tau_i > 0.5$. Since by definition we have that $\tau_i \equiv F_{t,t-h}(\widetilde{f}_{i,t}^h)$, and from A4, $F_{t,t-h}$ is a continuous function, it follows that when $\beta_1 = \beta_2$ we must also have that $\tau_i = 0.5 \therefore \widetilde{f}_{i,t}^h = Med_{t-h}(y_t)$. ■

Proof of Corollary 3. From Proposition 2, we have that (symmetric case) $\tau_i = 0.5$. In addition, if the conditional density function (pdf) of y_t is a unimodal, continuous and symmetric function about its conditional mean, then we also have that the conditional mean of y_t must coincide with the respective conditional median. In other words, it follows that $\widetilde{f}_{i,t}^h = F_{y_t}^{-1}(0.5 \mid \mathcal{F}_{t-h}) = Med_{t-h}(y_t) = \mathbb{E}_{t-h}(y_t)$. ■

Proof of Corollary 4. Part of this proof comes from Granger and Newbold (1986, p.126). Firstly, note (from A3) that the lin-lin loss function can be represented by $L(e) = g(e)h(e)$, in which $g(e) = \begin{cases} -e \ ; \ e < 0 \\ e \ ; \ e > 0 \end{cases}$ and $g'(e) = \begin{cases} -1 \ ; \ e < 0 \\ 1 \ ; \ e > 0 \end{cases}$.

Thus, from the proof of Proposition 1, we have that $\gamma_1 \equiv \int\limits_{-\infty}^{\widetilde{y}^i} g'(y - \widetilde{y}^i)f(y)dy = -\int\limits_{-\infty}^{\widetilde{y}^i} f(y)dy = -F_{t,t-h}(\widetilde{y}^i) = -\tau_i$; and $\gamma_2 \equiv \int\limits_{\widetilde{y}^i}^{\infty} g'(y - \widetilde{y}^i)f(y)dy = \int\limits_{\widetilde{y}^i}^{\infty} f(y)dy =$

$1-\int\limits_{-\infty}^{\widetilde{y}^i} f(y)dy = 1-F_{t,t-h}(\widetilde{y}^i) = 1-\tau_i$. In addition, we also have that $\beta_1\gamma_1 = -\beta_2\gamma_2$ $\therefore -\beta_1\tau_i = -\beta_2(1-\tau_i) = -\beta_2+\beta_2\tau_i \therefore \beta_2 = (\beta_1+\beta_2)\tau_i \therefore \tau_i = \beta_2/(\beta_1+\beta_2)$. ■

Proof of Proposition 5. From Proposition 1, recall that under an asymmetric loss function ($\beta_1 \neq \beta_2$) it follows that the optimal forecast $\widetilde{f}^h_{i,t}$ is a conditional quantile of y_t other than the median, such that $\tau_i \equiv F_{t,t-h}(\widetilde{f}^h_{i,t})$.

(ia) Gaussian distribution (μ, σ^2): $\widetilde{f}^h_{i,t} = \mu_{t,t-h} + \sigma_{t,t-h}\Phi^{-1}(\tau_i) = \mathbb{E}_{t-h}(y_t) + k^h_i$, where $k^h_i = \sigma_{t,t-h}\Phi^{-1}(\tau_i)$. Notice that under symmetry ($\beta_1 = \beta_2$) it follows that $k^h_i = \sigma_{t,t-h}\Phi^{-1}(\tau_i = 0.5) = 0$.

(ib) Two-piece normal $(\mu, \sigma_1^2, \sigma_2^2)$: According to Julio (2007), the conditional quantiles of a two-piece normal distribution, where μ denotes $\mathbb{E}_{t-h}(y_t)$, are given by: $\mu - \sqrt{2/\pi}(\sigma_2-\sigma_1) + \sigma_1\Phi^{-1}(\frac{\tau}{C\sqrt{2\pi}\sigma_1})$ where $C = \sqrt{\frac{2}{\pi}}(\sigma_1+\sigma_2)^{-1}$ for $\tau \leq \Pr[x \leq \mu - \sqrt{2/\pi}(\sigma_2-\sigma_1)]$; or equal to $\mu - \sqrt{2/\pi}(\sigma_2-\sigma_1) + \sigma_2\Phi^{-1}(\frac{\tau + C\sqrt{2\pi}\sigma_2 - 1}{C\sqrt{2\pi}\sigma_2})$ for $\tau > \Pr[x \leq \mu - \sqrt{2/\pi}(\sigma_2-\sigma_1)]$. In both cases, notice that $\widetilde{f}^h_{i,t} = \mathbb{E}_{t-h}(y_t) + k^h_i$.

(ic) Logistic distribution (a, b): The conditional quantiles are given by $a + b\ln\left(\frac{\tau}{1-\tau}\right)$, where $a = \mathbb{E}_{t-h}(y_t) \therefore \widetilde{f}^h_{i,t} = \mathbb{E}_{t-h}(y_t) + k^h_i$, where $k^h_i = b\ln\left(\frac{\tau_i}{1-\tau_i}\right)$. Notice that under symmetry ($\beta_1 = \beta_2$) of the loss function it follows that $k^h_i = b\ln\left(\frac{0.5}{1-0.5}\right) = 0$.

(iia) Log-normal (μ, σ^2): The conditional quantiles are given by $\exp(\mu + \sigma\Phi^{-1}(\tau))$, thus, $\widetilde{f}^h_{i,t} = \exp(\mu_{t,t-h} + \sigma_{t,t-h}\Phi^{-1}(\tau_i))$, whereas $\mathbb{E}_{t-h}(y_t) = \exp(\mu_{t,t-h} + \sigma^2_{t,t-h}/2)$. Therefore, it follows that $\widetilde{f}^h_{i,t} = \beta^h_i\mathbb{E}_{t-h}(y_t)$, where $\beta^h_i = \exp(\sigma_{t,t-h}\Phi^{-1}(\tau_i) - \sigma^2_{t,t-h}/2)$.

(iib) Weibull (b, k): The conditional quantiles are given by $b\left[-\ln(1-\tau)\right]^{1/k}$, thus, $\widetilde{f}^h_{i,t} = b\left[-\ln(1-\tau_i)\right]^{1/k}$, whereas $\mathbb{E}_{t-h}(y_t) = b\Gamma(1+1/k)$. Therefore, it follows that $\widetilde{f}^h_{i,t} = \beta^h_i\mathbb{E}_{t-h}(y_t)$, where $\beta^h_i = \frac{[-\ln(1-\tau_i)]^{1/k}}{\Gamma(1+1/k)}$.

(iii) Beta $(a = 1, b > 0)$: The conditional quantiles are given by $1-(1-\tau)^{1/b}$, whereas $\mathbb{E}_{t-h}(y_t) = \frac{a}{a+b} = \frac{1}{1+b}$, then, $\widetilde{f}^h_{i,t} = k^h_i + \beta^h_i\mathbb{E}_{t-h}(y_t)$; where $k^h_i = -(1-\tau_i)^{1/b}$ and $\beta^h_i = 1+b$.

(iv) Beta $(a > 0, b = 1)$: The conditional quantiles are given by $\tau^{1/a}$, then, $\ln(\widetilde{f}^h_{i,t}) = \ln(\tau_i^{1/a}) \therefore a = \frac{\ln(\tau_i)}{\ln(\widetilde{f}^h_{i,t})}$. Since $\mathbb{E}_{t-h}(y_t) = \frac{a}{a+b} = \frac{a}{a+1} = \frac{\frac{\ln(\tau_i)}{\ln(\widetilde{f}^h_{i,t})}}{1+\frac{\ln(\tau_i)}{\ln(\widetilde{f}^h_{i,t})}} = \frac{\ln(\tau_i)}{\ln(\tau_i)+\ln(\widetilde{f}^h_{i,t})} \therefore \ln(\widetilde{f}^h_{i,t}) = \frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)} - \ln(\tau_i) \therefore \widetilde{f}^h_{i,t} = \exp(\frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)})\exp(-\ln(\tau_i))$. Therefore, $\widetilde{f}^h_{i,t} = \beta^h_i\xi(\mathbb{E}_{t-h}(y_t))$, in which $\beta^h_i = \exp(-\ln(\tau_i))$ and ξ is the non-linear function $\xi(\mathbb{E}_{t-h}(y_t)) = \exp(\frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)})$. ■

Proof of Proposition 6. Parts of this proof follow Granger (1969), Patton and Timmermann (2007) and Gaglianone and Lima (2012). Recall that, without loss of generality, we assumed that $X'_{t,t-h} = (1, x_{t,t-h})$ is a 2×1 vector and $\delta = (\delta_0, \delta_1)'$; $\gamma = (\gamma_0, \gamma_1)'$. By homogeneity of the loss function and DGP (A5), the optimal forecast can be represented in the following way:

$$\begin{aligned}
\widetilde{f}^h_{i,t} &= \underset{\widetilde{y}}{\arg\ \min} \int L^i(y-\widetilde{y})dF_{t,t-h}(y) \\
& \underset{\widetilde{y}}{\arg\ \min} \int \left[g\left(\frac{1}{X'_{t,t-h}\gamma}\right)\right]^{-1} L^i\left(\frac{1}{X'_{t,t-h}\gamma}(y-\widetilde{y})\right) dF_{t,t-h}(y) \\
&= \underset{\widetilde{y}}{\arg\ \min} \int \left[g\left(\frac{1}{(\gamma_0+\gamma_1 x_{t,t-h})}\right)\right]^{-1} L^i\left(\frac{1}{(\gamma_0+\gamma_1 x_{t,t-h})}(y-\widetilde{y})\right) dF_{t,t-h}(y) \\
&= \underset{\widetilde{y}}{\arg\ \min} \int L^i\left(\frac{1}{(\gamma_0+\gamma_1 x_{t,t-h})}(y-\widetilde{y})\right) dF_{t,t-h}(y) \\
&= \underset{\widetilde{y}}{\arg\ \min} \int L^i\left(\frac{1}{(\gamma_0+\gamma_1 x_{t,t-h})}(\delta_0+\delta_1 x_{t,t-h}+\gamma_0\eta_t+\gamma_1 x_{t,t-h}\eta_t-\widetilde{y})\right) dF_{\eta,h}(\eta).
\end{aligned}$$

Now represent a forecast $\widetilde{y}$ of y_t made at period $t-h$ by $\delta_0+\delta_1 x_{t,t-h}+(\gamma_0+\gamma_1 x_{t,t-h})\,\widetilde{\gamma}$. This way, the optimal forecast $\widetilde{f}^h_{i,t}$ is given by:

$$\begin{aligned}
\widetilde{f}^h_{i,t} &= \delta_0+\delta_1 x_{t,t-h}+(\gamma_0+\gamma_1 x_{t,t-h})\cdot \underset{\widetilde{\gamma}}{\arg\ \min} \int L^i\left(\frac{1}{(\gamma_0+\gamma_1 x_{t,t-h})}(\delta_0+\delta_1 x_{t,t-h}\right. \\
&\quad \left. +(\gamma_0+\gamma_1 x_{t,t-h})\,\eta_t-\delta_0-\delta_1 x_{t,t-h}-(\gamma_0+\gamma_1 x_{t,t-h})\,\widetilde{\gamma})\right)dF_{\eta,h}(\eta) \\
&= \delta_0+\delta_1 x_{t,t-h}+(\gamma_0+\gamma_1 x_{t,t-h})\cdot \underset{\widetilde{\gamma}}{\arg\ \min} \int L^i\left(\eta_t-\widetilde{\gamma}\right) dF_{\eta,h}(\eta) \\
&= \delta_0+\gamma_0\gamma^i_h+\delta_1 x_{t,t-h}+\gamma_1 x_{t,t-h}\gamma^i_h \\
&= \alpha_0(\tau_i)+\alpha_1(\tau_i)x_{t,t-h} \text{ where } \alpha_0(\tau_i) = \left(\delta_0+\gamma_0\gamma^i_h\right) \text{ and } \alpha_1(\tau_i) = (\delta_1+\gamma_1\gamma^i_h),
\end{aligned}$$

in which we have used the fact that $F_{\eta,h}(\eta)$ is time-invariant by definition, and $\gamma^i_h \equiv \underset{\widetilde{\gamma}}{\arg\ \min} \int L^i\left(\eta_t-\widetilde{\gamma}\right) dF_{\eta,h}$ or, equivalently, $\gamma^i_h = F^{-1}_{\eta,h}(\tau_i)$ since $\widetilde{f}^h_{i,t} = F^{-1}_{t,t-h}(\tau_i)$. Therefore, the location-scale assumption for the DGP implies that the optimal forecast can be viewed as a linear conditional quantile of y_t, evaluated at the specific percentile $\tau_i \in [0,1]$, so that $\widetilde{f}^h_{i,t} = \alpha_0(\tau_i)+\alpha_1(\tau_i)x_{t,t-h}$.

On the other hand, from Koenker (2005, p.302), we know that integrating a conditional quantile function of y_t over the entire domain $\tau \in [0,1]$ yields the conditional mean of y_t. In other words, provided that y_t is given by a location-scale model (A5), it follows that a conditional quantile of y_t is given by $F^{-1}_{t,t-h}(\tau) = \alpha_0(\tau) + \alpha_1(\tau)x_{t,t-h}$, for some $\tau \in [0,1]$ $\therefore$ $\mathbb{E}_{t-h}(y_t) = \int_0^1 F^{-1}_{t,t-h}(\tau)d\tau = \int_0^1 (\alpha_0(\tau) + \alpha_1(\tau)x_{t,t-h})\, d\tau = \overline{\alpha_0} + \overline{\alpha_1}x_{t,t-h}$, where $\overline{\alpha_j} = \int_0^1 \alpha_j(\tau)d\tau$; $j = \{0;1\}$. Thus, $\mathbb{E}_{t-h}(y_t) = \overline{\alpha_0} + \overline{\alpha_1}x_{t,t-h}$ $\therefore$ $\frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}\mathbb{E}_{t-h}(y_t) = \overline{\alpha_0}\frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} + \overline{\alpha_1}\frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}x_{t,t-h}$ $\therefore$ $\frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}\mathbb{E}_{t-h}(y_t) + \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}}\alpha_1(\tau_i)\right) = \overline{\alpha_0}\frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} + \overline{\alpha_1}\frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}x_{t,t-h} + \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}}\alpha_1(\tau_i)\right) = \alpha_0(\tau_i) + \alpha_1(\tau_i)x_{t,t-h} = \widetilde{f}^h_{i,t}$. Therefore, if one defines $k^h_i \equiv \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}}\alpha_1(\tau_i)\right)$ and $\beta^h_i \equiv \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}$ it follows that $\widetilde{f}^h_{i,t} = k^h_i + \beta^h_i\mathbb{E}_{t-h}(y_t)$. Notice that k^h_i and β^h_i are functions of $\alpha_0(\tau)$ and $\alpha_1(\tau)$, which depend on the parameters δ and γ of the location-scale model and on γ^i_h, which is a constant that depends only on the distribution $F_{\eta,h}(0,1)$ and the loss function L^i.

In the case of no scale effects on the DGP, it follows that only the intercept function $\alpha_0(\tau)$ varies across the quantile levels τ and, thus, it follows that $\alpha_1(\tau) = \overline{\alpha_1}$ for all $\tau \in [0,1]$ $\therefore$ $\beta^h_i = 1$. ■

Proof of Proposition 7. (i) From Proposition 6, it follows that $\widetilde{f}^h_{i,t} = k^h_i + \beta^h_i\mathbb{E}_{t-h}(y_t)$, where $k^h_i \equiv \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}}\alpha_1(\tau_i)\right)$ and $\beta^h_i \equiv \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}$; $\alpha_0(\tau_i) = (\delta_0 + \gamma_0\gamma^i_h)$; $\alpha_1(\tau_i) = (\delta_1 + \gamma_1\gamma^i_h)$; $\overline{\alpha_j} = \int_0^1 \alpha_j(\tau)d\tau$ for $j = \{0;1\}$. In addition, the location-scale model implies that $\mathbb{E}_{t-h}(y_t) = \overline{\alpha_0} + \overline{\alpha_1}x_{t,t-h}$. Notice that k^h_i and β^h_i are functions of $\alpha_0(\tau_i)$ and $\alpha_1(\tau_i)$, which (in turn) depend on the parameters δ and γ of the location-scale model and on γ^i_h, which is a constant that only depends on the distribution $F_{\eta,h}(0,1)$ and on the loss L^i.

Nonetheless, the optimal (feasible) forecast $f^h_{i,t}$ of y_t conditioned on the information set available at period $(t-h)$ is given by $f^h_{i,t} = \widehat{k}^h_i + \widehat{\beta}^h_i\widehat{\mathbb{E}}_{t-h}(y_t)$. If we define the error term $\varepsilon^h_{i,t}$ as the difference between the optimal (feasible) forecast and the optimal forecast, it follows that $\varepsilon^h_{i,t} \equiv f^h_{i,t} - \widetilde{f}^h_{i,t} = f^h_{i,t} - k^h_i - \beta^h_i\mathbb{E}_{t-h}(y_t)$ $\therefore$ $f^h_{i,t} = k^h_i + \beta^h_i\mathbb{E}_{t-h}(y_t) + \varepsilon^h_{i,t}$, where $\varepsilon^h_{i,t} \neq 0$, provided that $\widehat{\alpha}_0(\tau) - \alpha_0(\tau) \neq 0$ and $\widehat{\alpha}_1(\tau) - \alpha_1(\tau) \neq 0$, for all $\tau \in [0;1]$, due to parameter uncertainty, under a finite sample with T observations of $\{y_t; x_{t,t-h}\}_{t=1}^T$.

(ii) First define $\widehat{\alpha}(\tau) \equiv [\widehat{\alpha_0}(\tau); \widehat{\alpha_1}(\tau)]$. Koenker (2005, Theorem 4.1, p.120) shows that the estimator $\widehat{\alpha}_n(\tau) = \underset{\alpha\in\mathbb{R}^2}{\arg\min} \sum_{t=1}^n \rho_\tau(y_t - \alpha_0 - \alpha_1 x_{t,t-h})$, where ρ_τ

is defined as in Koenker and Basset (1978) by $\rho_{\tau}(u) = \begin{cases} \tau u, u \geq 0 \\ (\tau - 1) u, u < 0 \end{cases}$, is a consistent estimator for $\alpha(\tau)$, in linear conditional quantile functions (under some mild regularity conditions on $x_{t,t-h}$), as long as $n \to \infty$. This way, it follows that $\widehat{\overline{\alpha}} = \frac{1}{K}\sum_{k=1}^{K}\widehat{\alpha}(\tau_k)\Delta\tau_k \xrightarrow{p} \overline{\alpha} = \int_0^1 \alpha(\tau)d\tau$, provided that $\widehat{\alpha}(\tau) \xrightarrow{p} \alpha(\tau)$. Therefore, it also follows that $\widehat{k}_i^h \equiv \left(\widehat{\alpha}_0(\tau_i) - \frac{\widehat{\overline{\alpha_0}}}{\widehat{\overline{\alpha_1}}}\widehat{\alpha}_1(\tau_i)\right) \xrightarrow{p} k_i^h = \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}}\alpha_1(\tau_i)\right)$ and $\widehat{\beta}_i^h \equiv \frac{\widehat{\alpha}_1(\tau_i)}{\widehat{\overline{\alpha_1}}} \xrightarrow{p} \beta_i^h = \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}$.

(iii) From item (ii), it follows that $[\widehat{k}_i^h; \widehat{\beta}_i^h] \xrightarrow{p} [k_i^h; \beta_i^h]$. Thus, it follows that $\widehat{B^h} \equiv \frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i^h \xrightarrow{p} B^h$ and $\widehat{\beta}^h \equiv \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_i^h \xrightarrow{p} \beta^h$. The location-scale model also implies (see proof of Proposition 6) that $\mathbb{E}_{t-h}(y_t) = \overline{\alpha_0} + \overline{\alpha_1}x_{t,t-h}$. If one defines the sample analog $\widehat{\mathbb{E}}_{t-h}(y_t) = \widehat{\overline{\alpha_0}} + \widehat{\overline{\alpha_1}}x_{t,t-h}$, from consistency of $[\widehat{\alpha_0}(\tau); \widehat{\alpha_1}(\tau)]$ and provided that $x_{t,t-h}$ is covariance-stationary, it follows that $\widehat{\mathbb{E}}_{t-h}(y_t) \xrightarrow{p} \mathbb{E}_{t-h}(y_t)$. Now consider the (feasible) optimal forecast $f_{i,t}^h \equiv \widehat{k}_i^h + \widehat{\beta}_i^h\widehat{\mathbb{E}}_{t-h}(y_t)$. Taking the cross-sectional average, it follows that $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{k}_i^h + \widehat{\beta}_i^h\widehat{\mathbb{E}}_{t-h}(y_t)\right) = \frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i^h + \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\beta}_i^h\widehat{\mathbb{E}}_{t-h}(y_t)\right) = \widehat{B^h} + \widehat{\beta^h}\widehat{\mathbb{E}}_{t-h}(y_t) \therefore \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}} = \widehat{\mathbb{E}}_{t-h}(y_t)$. Now, taking the sequential limits on T and N, it follows that $\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}}\right) = \mathbb{E}_{t-h}(y_t)$.
■

Proof of Proposition 8. Let first $T \to \infty$ to obtain $\underset{T\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}}\right) = \underset{T\to\infty}{\text{plim}}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i^h}{\frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_i^h}\right) = \frac{\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i^h}{\frac{1}{N}\sum_{i=1}^{N}\beta_i^h} = y_t + \eta_t^h + \frac{\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h}{\frac{1}{N}\sum_{i=1}^{N}\beta_i^h}$, where the second equality comes from A12 (consistency of GMM estimates, e.g. under Theorem 2.6 of Newey and McFadden, 1994). Now, letting $N \to \infty$ we obtain (by A10 and Lemma 1 of Issler and Lima, 2009) $\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h\right) = 0$ and (by A7, A10, A11) that $\underset{N\to\infty}{\text{plim}}\left(\frac{\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h}{\frac{1}{N}\sum_{i=1}^{N}\beta_i^h}\right) = 0 \therefore \underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}}\right) = y_t + \eta_t^h = \mathbb{E}_{t-h}(y_t)$ ■

A.2 Parameter uncertainty: Some examples

Example 1 Suppose that: (i) DGP is given by $\left(y_{t|t-h}\right) \sim N(\mu, \sigma^2)$; (ii) Loss function of individual i is the MSE function; (iii) Individual i (correctly) assumes that the DGP is given by a Gaussian distribution, however, due to finite sample, estimates the parameters $(\widehat{\mu}, \widehat{\sigma}^2)$. In this case, from (i) and (ii), the optimal forecast (Granger, 1969) is $\widetilde{f}_{i,t}^h = E_{t-h}(y_t) = \mu$. However, the best that individual i can do is to approximate the "optimal forecast" by using (ii) and (iii), so that $f_{i,t}^h = \widehat{\mathbb{E}}_{t-h}(y_t) = \widehat{\mu}$. Thus, the optimal forecast "approximation error" is given by $\varepsilon_{i,t}^h \equiv f_{i,t}^h - \widetilde{f}_{i,t}^h = \widehat{\mu} - \mu$. Provided that $\widehat{\mu}$ is consistently estimated, it follows that $\underset{T\to\infty}{\text{plim}} \left(\varepsilon_{i,t}^h\right) = 0$.

Example 2 Suppose that: (i) DGP is given by $\left(y_{t|t-h}\right) \sim N(\mu, \sigma^2)$; (ii) Loss function of individual i is the Lin-Lin asymmetric function of Corollary 4; (iii) Individual i (correctly) assumes that the DGP is given by a Gaussian distribution, however, due to finite sample, estimates the parameters $(\widehat{\mu}, \widehat{\sigma}^2)$. In this case, from (i) and (ii), the optimal forecast (see Proposition 5 and Corollary 4) is $\widetilde{f}_{i,t}^h = k_i^h + E_{t-h}(y_t)$, where $k_i^h = \sigma\Phi^{-1}(\tau_i)$ and $\tau_i = \beta_2/(\beta_1 + \beta_2)$. However, individual i approximates the "optimal forecast" by using (ii) and (iii), so that $f_{i,t}^h = \widehat{\mu} + \widehat{\sigma}\Phi^{-1}(\tau_i)$. Thus, the optimal forecast "approximation error" is given by $\varepsilon_{i,t}^h \equiv f_{i,t}^h - \widetilde{f}_{i,t}^h = (\widehat{\mu} - \mu) + (\widehat{\sigma} - \sigma)\,\Phi^{-1}(\tau_i)$. Provided that $(\widehat{\mu}, \widehat{\sigma}^2)$ are consistently estimated, it follows that $\underset{T\to\infty}{\text{plim}} \left(\varepsilon_{i,t}^h\right) = 0$.

Example 3 Suppose that: (i) DGP is given by $\left(y_{t|t-h}\right) \sim Weibull(b, k)$; (ii) Loss function of individual i is the Lin-Lin asymmetric function of Corollary 4; (iii) Individual i (wrongly) assumes that the DGP is given by a Gaussian distribution, and estimate the parameters $(\widehat{\mu}, \widehat{\sigma}^2)$. In this case, from (i) and (ii), the optimal forecast (see proof of Proposition 5) is $\widetilde{f}_{i,t}^h = b\left[-\ln(1-\tau_i)\right]^{1/k}$, where $\tau_i = \beta_2/(\beta_1 + \beta_2)$. However, individual i try to approximate the "optimal forecast" by using (ii) and (iii), so that $f_{i,t}^h = \widehat{\mu} + \widehat{\sigma}\Phi^{-1}(\tau_i)$. This way, $\varepsilon_{i,t}^h \equiv f_{i,t}^h - \widetilde{f}_{i,t}^h = \widehat{\mu} + \widehat{\sigma}\Phi^{-1}(\tau_i) - b\left[-\ln(1 - \beta_2/(\beta_1 + \beta_2))\right]^{1/k}$ and, thus, $\underset{T\to\infty}{\text{plim}} \left(\varepsilon_{i,t}^h\right) \neq 0$.

A.3 Results of BCAF: $\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^{h} - \widehat{B^h} \right) = \mathbb{E}_{t-h}\left(y_t\right)$

Figure 1 - CPI inflation rate in Brazil (y_t)
Monthly and twelve-month accumulated rates (%)

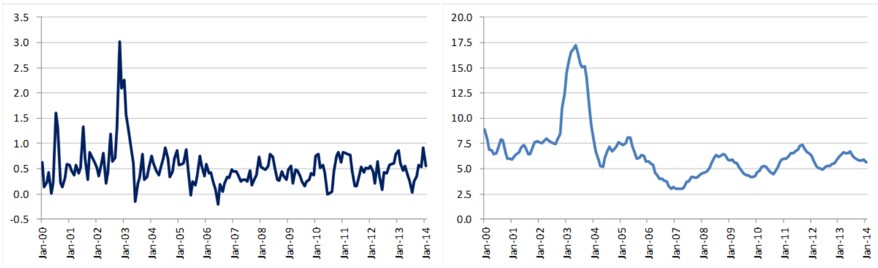
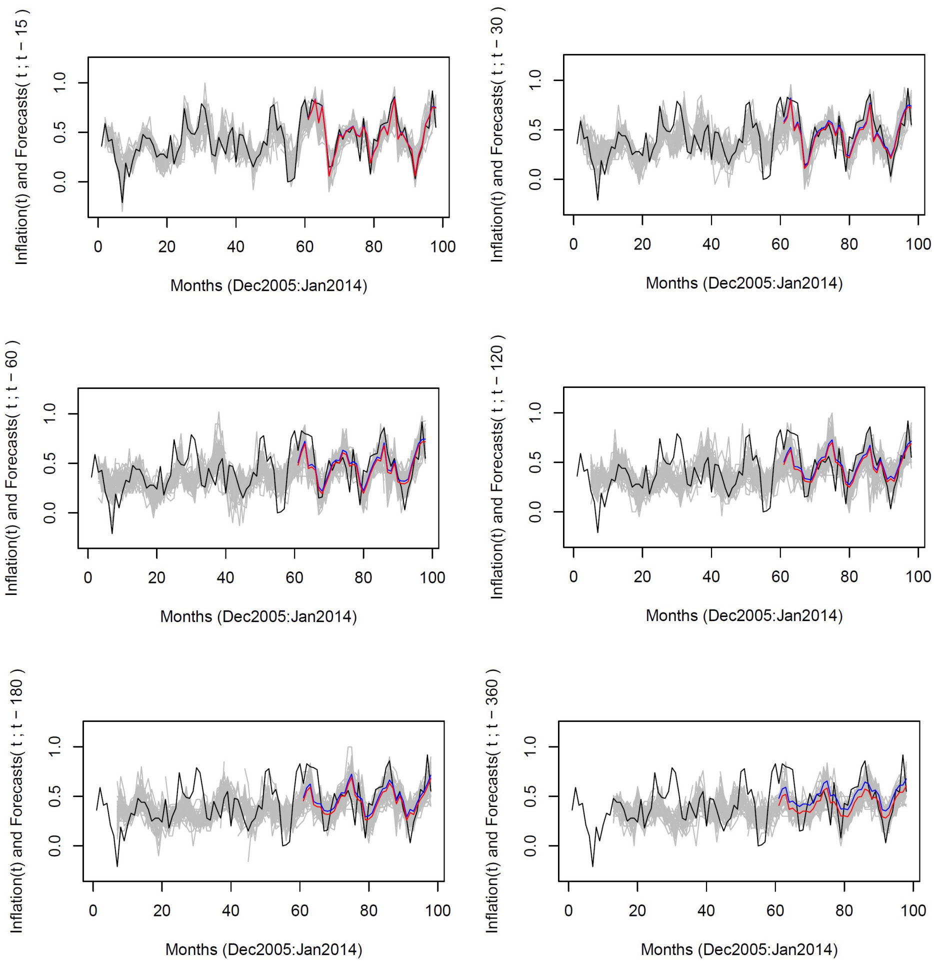


Figure 2 - Inflation rate (y_t) and survey-based forecasts ($f_{i,t}^h$)
for selected horizons (in days)



Note: The black line shows the monthly inflation rate. The red line represents the average forecast and the blue line shows the BCAF forecast. Gray lines show the forecasts $f_{i,t}^h$ of survey participant i for y_t made at period $t-h$.

Figure 3 - Number of agents (N) for each forecast horizon (h)

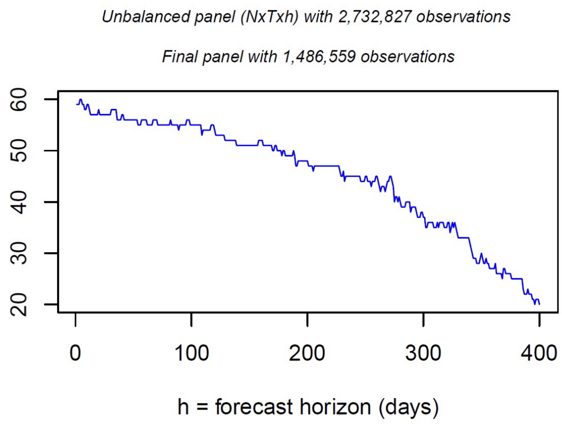


Table 1: Average Bias - BCAF (Issler and Lima, 2009)

Forecast Horizon h (in months)	Average Bias $\widehat{B^h}$	$H_0 : B^h = 0$ p-value
1	-0.0188	0.293
2	-0.0259	0.255
3	-0.0261	0.270
6	-0.0324	0.198
9	-0.0568	0.004
12	-0.0715	0.000

Table 2: MSE comparison

Forecast Horizon h (in months)	(a) MSE Average Forecast	(b) MSE BCAF	(a)/(b)
1	0.0177	0.0166	1.06
2	0.0248	0.0234	1.06
3	0.0267	0.0254	1.05
6	0.0312	0.0294	1.06
9	0.0328	0.0293	1.12
12	0.0357	0.0309	1.16

Figure 4 - Forecast Error Term-Structure ($\widehat{B^h}$)

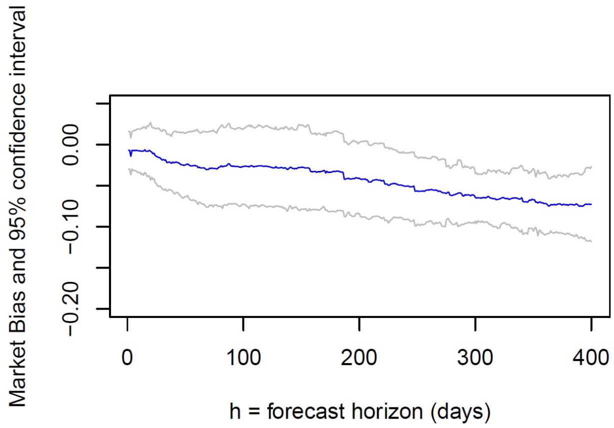
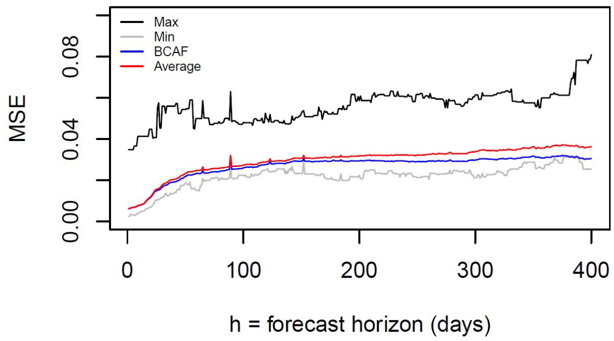


Figure 5 - Mean Squared Error (MSE)



Note: Max and Min denote the maximum and minimum MSEs, for each horizon, across all forecasters. Average refers to the MSE of the average forecast.

Figure 6 - Relative MSE

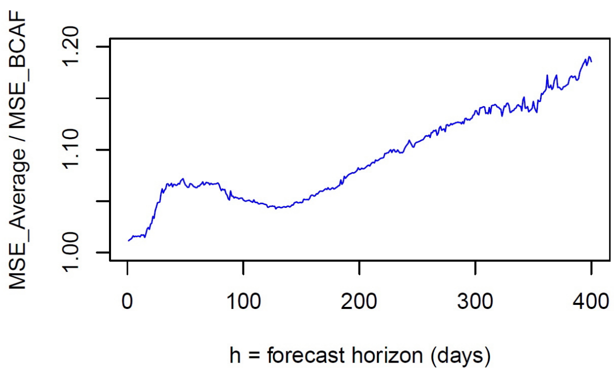
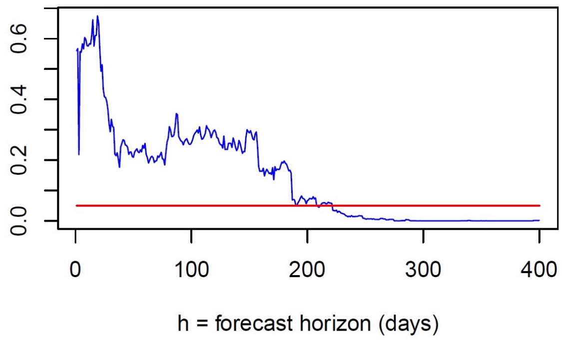


Figure 7 - Hypothesis test of null-bias ($Ho : B^h = 0$); p-values of a t-ratio test



Note: Red line shows a p-value of 0.05.

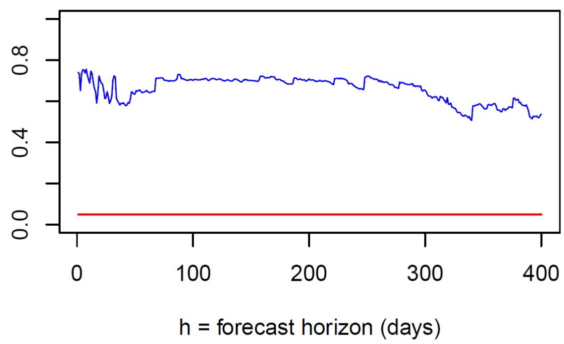
A.4 Extended BCAF: $\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{B^h}}{\widehat{\beta^h}} \right) = \mathbb{E}_{t-h}(y_t)$

Table 3 - Estimation results

	Model 1: Disaggregated Forecasts			Model 2: Aggregated Forecasts		
horizon (days)	$\hat{B}$ (h)	$\hat{\text{beta}}$ (h)	Wald test (p-value)	$\hat{B}$ (h)	$\hat{\text{beta}}$ (h)	Wald test (p-value)
10	0.0053	0.9673	2.4E-92	0.0036	0.9638	1.1E-13
	(0.0003)	(0.0063)		(0.0009)	(0.0276)	
20	0.0098	0.9308	5.7E-124	0.0071	0.9312	3.0E-12
	(0.0005)	(0.0074)		(0.0016)	(0.0399)	
30	0.0147	0.8858	5.9E-233	0.0118	0.8929	7.5E-16
	(0.0006)	(0.0086)		(0.0026)	(0.0481)	
60	0.0205	0.8620	4.1E-149	0.0173	0.8541	9.6E-11
	(0.0009)	(0.0147)		(0.0033)	(0.0583)	
90	0.0223	0.8500	1.1E-226	0.0180	0.8647	1.8E-12
	(0.0009)	(0.0169)		(0.003)	(0.0763)	
180	0.0217	0.8669	8.1E-188	0.0176	0.8486	2.8E-11
	(0.0009)	(0.0211)		(0.003)	(0.0884)	
360	0.0236	0.7965	4.7E-112	0.0199	0.8275	2.9E-15
	(0.0013)	(0.0162)		(0.0025)	(0.088)	

Note: Standard deviation in parentheses. Wald test refers to Ho: [B(h);beta(h)] = [0;1].

Figure 8 - GMM overidentification restrictions; TJ test (p-values)



Note: Red line shows a p-value of 0.05.

Figure 9 - Coefficient $\widehat{B^h}$ and 95% confidence interval

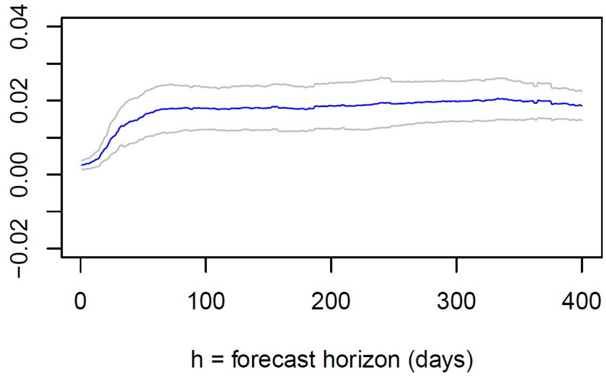


Figure 10 - Coefficient $\widehat{\beta^h}$ and 95% confidence interval

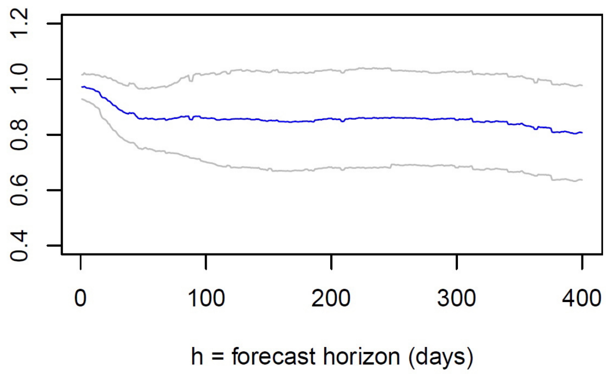
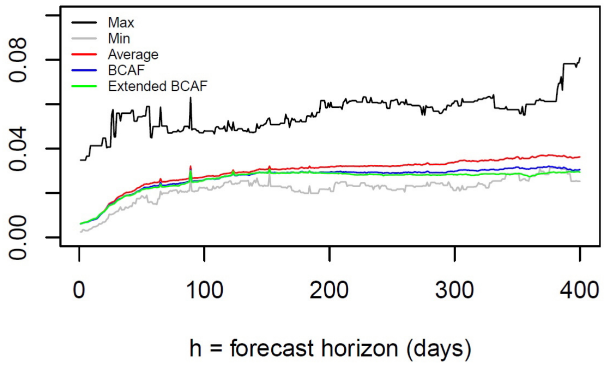


Figure 11 - MSE comparison



Note: Max and Min denote the maximum and minimum MSEs, for each horizon, across all forecasters. Average refers to the MSE of the average forecast.

Figure 12 - MSE ratio comparison

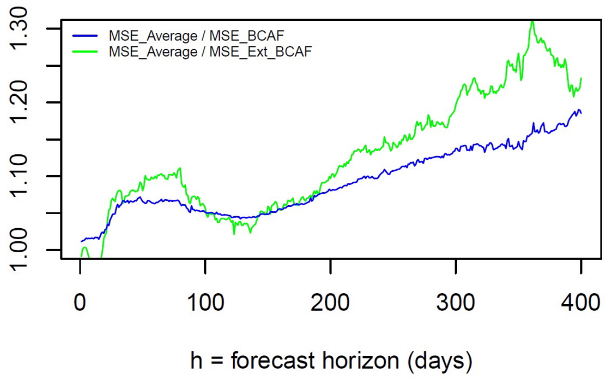
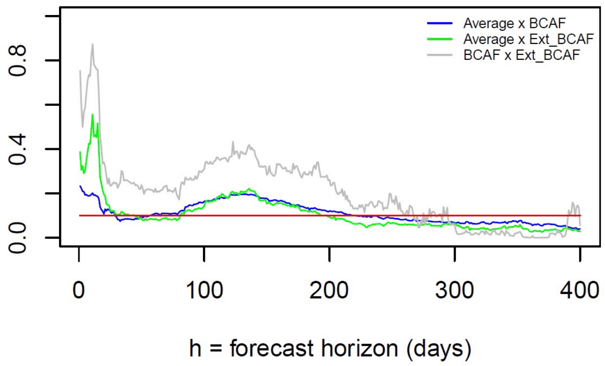


Figure 13 - Clark and West (2007) test (p-values)
Ho: equal predictive accuracy



Note: Red line shows a p-value of 0.10.

Figure 14 - Aggregate shock $\widehat{\eta_t^h}$ for selected horizons (in months)

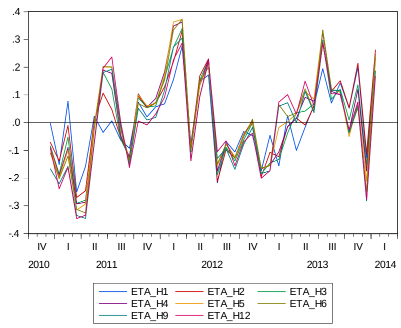


Table 4: Optimality test: η_t^h follows a MA of order (at most) $h-1$

Based on the regression: $\eta_t^h = a + \epsilon_t^h + \sum_{j=1}^{q+h} b_j \epsilon_{t-j}^h$

Forecast Horizon h (in months)	$H_0 : a = b_j = 0;$ for all $j \geq h$ (p-value)	
	$q = 2$	$q = 3$
1	0.003	0.000
2	0.534	0.559
3	0.828	0.612
6	0.392	0.049
9	0.011	0.007
12	0.024	0.412

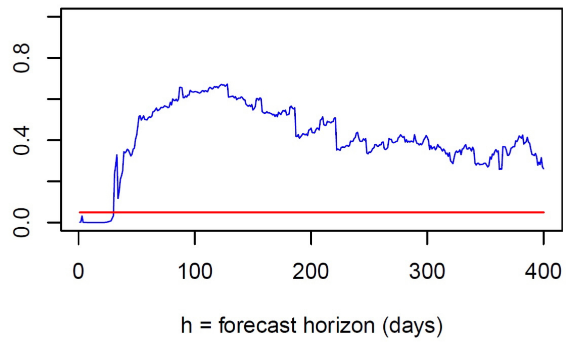
A.5 Aggregate Forecasts and Bias-Correction (NLS x GMM)

Table 5 - Estimation comparison

	Nonlinear Least Squares		GMM estimation	
horizon (days)	$\hat{B}$ (h)	$\hat{\text{beta}}$ (h)	$\hat{B}$ (h)	$\hat{\text{beta}}$ (h)
10	0.0547	0.8372	0.0036	0.9638
	(0.0185)	(0.0429)	(0.0009)	(0.0276)
20	0.0919	0.7376	0.0071	0.9312
	(0.0268)	(0.0635)	(0.0016)	(0.0399)
30	0.0746	0.7536	0.0118	0.8929
	(0.0441)	(0.1078)	(0.0026)	(0.0481)
60	-0.0250	0.9880	0.0173	0.8541
	(0.1105)	(0.2837)	(0.0033)	(0.0583)
90	-0.0392	1.0306	0.0180	0.8647
	(0.1369)	(0.3542)	(0.003)	(0.0763)
180	-0.0672	1.0834	0.0176	0.8486
	(0.2045)	(0.5238)	(0.003)	(0.0884)
360	-0.3765	1.7374	0.0199	0.8275
	(0.6405)	(1.5355)	(0.0025)	(0.088)

Note: Standard deviation in parentheses. NLS coefficients from regression $y_t = -B^h/\beta^h + \overline{f}_t^h/\beta^h + \epsilon_t^h$, where $\overline{f}_t^h$ denotes the average forecast. GMM estimation based on Model 2.

Figure 15 - Nonlinear Least Squares (NLS) - Hypothesis test of null-bias (p-value) $Ho : (B^h = 0; \beta^h = 1)$



Note: Red line shows a p-value of 0.05.

Figure 16 - MSE ratio comparison

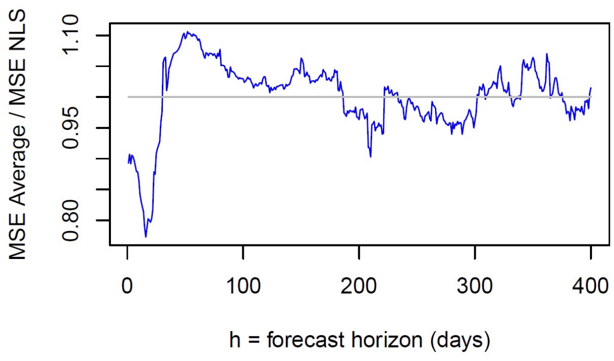


Figure 17 - MSE ratio comparison (cont.)

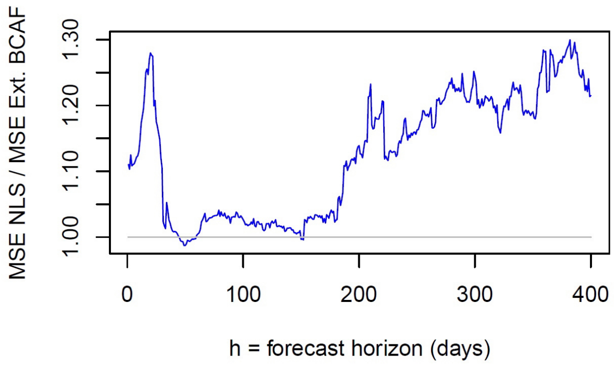
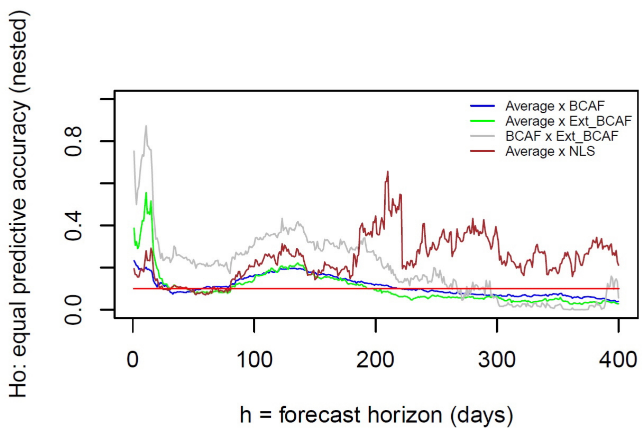


Figure 18 - Clark and West (2007) test (p-values)
Ho: equal predictive accuracy



Note: Red line shows a p-value of 0.10.