Microfounded Forecasting*

Wagner Piazza Gaglianone†

João Victor Issler‡

February 8, 2019

Abstract

This paper proposes a *financial approach* to economic forecasting which can be applied to data bases of surveys of forecasts. We model the forecasting decision of an individual from *first principles* (i.e., *microfounded*) and show that surveys of forecasts obey an affine factor structure with a single factor which is the common component of the conditional expectation of the target variable. This holds in a context where individuals only have access to public information or where they also have access to private information with common and idiosyncratic components. We show that asymptotically efficient forecasts of the target variable can be built using the generalized method of moments in a panel-data context, when N and T diverge or when T diverges with N fixed. In this context, the feasible optimal forecast is a function of the consensus forecast of the survey (a cross-sectional average of survey forecasts) after appropriately filtering out two bias terms. This links the *financial approach* of economic forecasting to the forecast-combination literature, where idiosyncratic risk of individual forecasts can be diversified out. Our microfounded approach is applied to a world-class data base on surveys of expectations and the techniques advanced here fare best when compared with competitive alternatives.

---

*The views expressed in the paper are those of the authors and do not necessarily reflect those of the Banco Central do Brasil or of FGV. We are especially grateful for the comments and suggestions given by Scott Atkinson, Francis Diebold, Robert Engle, Antonio Galvão, Raffaella Giacomini, Alain Hecq, Bo Honoré, Fabian Krüger, Michael McCracken, Marcelo Moreira, Whitney Newey, Andrew Patton, Barbara Rossi, Rafael Santos, Allan Timmermann, Mark Watson and Jonathan Wright. We also benefited from comments given by the seminar participants of the International Symposium on Forecasting (Seoul); International Panel Data Conference (London); Latin American Workshop in Econometrics (São Paulo); XVI Annual Inflation Targeting Seminar of the Banco Central do Brasil (Rio de Janeiro); The Conference of the International Association for Applied Econometrics (London); and the NBER Conference SI 2018 - Forecasting & Empirical Methods (Boston). Both Gaglianone and Issler gratefully acknowledge the support from CNPq, FAPERJ, INCT and FGV on different grants.

†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

Keywords: Forecast Combination, Affine Model, Common Factors (Features), Big Data.

JEL Classification: C14, C33, E37.

1 Introduction

The objective of this paper is to propose a *financial approach* to economic forecasting based on first principles. Our focus is on surveys of forecasts. We take seriously the optimal forecast choice faced by economic agents, where the minimized loss function can be of unknown form and the forecasting agent has no knowledge of the *data generating process* (DGP) of the target variable. In this context, we show that optimal forecasts obey a factor model with an affine structure, mimicking canonical models in finance. This allows employing this setup as a basis to forecast economic series combining point forecasts contained in the survey. As far as we know, this is the first paper to propose such a holistic approach connecting optimal forecasts and financial models.

Our first result shows that, if all agents possess the same information set consisting of public information, forecasts will obey an affine model where its common factor is the conditional expectation of the target variable based on the common information set. Using standard econometric techniques based on the generalized method of moments (GMM), we show how one can identify and estimate the common factor, i.e., the conditional expectation based on public information. An alternative identification and estimation method based on principal-component analysis is also feasible. In our second result, we extend the basic setup to deal with heterogeneous information sets consisting of public and private information. We show how one can recover the common component of information sets encompassing public and private information. In both cases, using a quadratic loss in a *pseudo*-out-of-sample forecasting exercise, our estimate of the common component performs best when compared with commonly used alternatives.

The motivation to propose a *financial approach* to economic forecasting is not new. The main idea behind forecast combinations is that, if the number of forecasts is large enough, an appropriate combination of these forecasts will eliminate idiosyncratic forecast errors, thereby reducing the uncertainty in forecasting. In 2006, Timmermann coined the phrase "*a financial approach to economic forecasting*" based on the idea of forecast combination originally proposed by Bates and Granger (1969): "A simple portfolio diversification argument motivates the idea of combining forecasts, c.f., Bates and Granger." Several authors have worked on what now could be labelled as a *financial approach* to economic forecasting, e.g., Granger and Ramanathan (1984), Palm and Zellner (1992), Stock and Watson (1999, 2002a and b), Elliott and Timmermann (2004), Capistrán and Timmermann (2009), Issler and Lima (2009), Gaglianone and Lima (2012, 2014), Hsiao and Wan (2014), and Lahiri, Peng, and Sheng (2015).

We propose the use of survey-based forecasts because they have outperformed model-based forecast in several contexts. Ang et al. (2007) argue that economists use four main methods to forecast inflation: time-series models of the ARIMA variety; structural models built upon the Phillips Curve; methods using information embedded in asset prices – in particular, the term-structure of interest rates; and methods that employ survey-based measures provided by economic agents (consumers and/or professional forecasters). They find that true out-of-sample survey forecasts (e.g. Michigan, Livingston) outperform a large number of out-of-sample single-equation and multivariate time-series competitors.

Faust and Wright (2013) argue that subjective forecasts of inflation seem to outperform model-based forecasts in certain dimensions, often by a wide margin. They discuss some reasons why this is the case, which points 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. One important observation of their analysis is that econometric models and surveys of forecasts can be thought as complementary. So focusing on the latter does not rule out using the former.

To handle survey of expectations data, we propose two layers of decision making. In the first layer, using standard microeconomic and econometric techniques, we first ask what is the optimal survey response from the point-of-view of a given respondent. As final users of that information, it is important to consider which environment we are dealing with. Indeed, it is best described as an environment where we do not know which loss function is being optimized by each agent posting forecasts and where agents have no knowledge of the DGP of the target variable. To make the optimization problem tractable, we assume that survey respondents employ the location-scale model to approximate the unknown DGP of the target variable. Here, heterogeneity of forecasts is naturally generated by agents employing different loss functions[1].

In the second layer of decision making, the final user of economic forecasts – *the econometrician* – observes several forecasts posted by economic agents, indexed by $i = 1, 2, \cdots, N$, across time, indexed by $t = 1, 2, \cdots, T$. If the final user of information employs a mean-squared-error (MSE) *risk function*, which is typically the case for the government, a central bank, a large risk-neutral firm, etc., it will be optimal to extract from the survey of expectations its common factor – the conditional expectation of the target variable[2] – which estimate will be a bias-corrected combination of forecasts.

[1]Potentially, it could also be due to agents employing different DGPs to forecast the target variable.

[2]Of course, different risk functions in the second layer would imply extracting a different target.

The techniques discussed above are appropriate to be applied to *current* surveys of economic forecasts, which are panels characterized by a limited number of respondents (limited N) and a large and increasing number of time observations (large T), e.g., The Survey of Professional Forecasters, The Michigan Survey, The Livingston Survey, The Bloomberg Survey of Economists, etc. In a world where there is an increasing availability of reliable data provided electronically, it is interesting to examine how one could efficiently use this wealth of information, since, in the future, we will be dealing more and more with *big data* surveys – very large N and T. The techniques proposed in this paper could be easily adapted to deal with future *big data* surveys as well. This is important since, as Einav and Levin (2014) have put it: "The most common uses of "big data" by companies are for tracking business processes and outcomes, and for building a wide array of predictive models".

The ideas in this paper are related to research done in three different fields. From econometrics, it is related to the common features literature after Engle and Kozicki (1993). Indeed, we attempt to bridge the gap between a large literature on common features applied to macroeconomics, e.g., Vahid and Engle (1993, 1997), Engle and Issler (1995), Issler and Vahid (2001, 2006) and Vahid and Issler (2002), Athanasopoulos et al. (2011), and the econometrics literature on forecasting related either to common factors or to forecast combination, perhaps best represented by the work of Bates and Granger (1969), Granger and Ramanathan (1984), Palm and Zellner (1992), Davies and Lahiri (1995), Forni et al. (2000, 2005), Stock and Watson (2002a,b, 2006), Elliott and Timmermann (2004, 2005), Timmermann (2006), Patton and Timmermann (2007), Issler and Lima (2009), Gaglianone et al. (2011), Gaglianone and Lima (2012, 2014), and Lahiri, Peng, and Sheng (2015). In this context, it is also related to the literature on the role of loss functions, e.g., Elliott et al. (2008) and Capistrán and Timmermann (2009). From *big data* econometrics it is related to the work of Einav and Levin (2014), Varian (2014), and Diebold (2012).

The techniques proposed here are applied to a world-class data base on inflation expectations – *The Focus Survey* of forecasts – a unique panel database put together by the Brazilian Central Bank (BCB). It collects *daily* expectations from professional forecasters in commercial banks, asset management firms, consulting firms and non-financial institutions, followed throughout time with a reasonable turnover. It contains a set of 254 registered institutions in the system, with a smaller *active* group of around 100 institutions. Our sample covers daily inflation forecasts of monthly IPCA inflation from January 2nd, 2006, to February 7th, 2014 ($2,020$ working days). The forecast horizon ranges from one day up to 400 days, forming an unbalanced panel containing a total amount of $2,732,827$ observations. Under MSE risk, our estimate of the common

This is not a problem here, given that we have properly characterized the first layer of decisions.

factor (feature) of inflation expectations fares better than frequently employed forecast-combination techniques, the *best ARMA* model chosen using Schwarz Criterion, as well as principal-component based techniques. When we compare our method with alternative forecast-combination techniques, MSE reductions can reach up to 25%. Reductions vis-à-vis the MSE associated with forecasts of the *best ARMA* model can reach up to 50%.

The rest of the paper is divided as follows. Section 2 introduces a microfounded-based framework to study the forecast error under risk functions more general than the usual MSE. Section 3 presents a real-time forecasting exercise with data from *The Focus Survey* of Brazilian inflation expectations, comparing the out-of-sample performance of different forecasting approaches. Section 4 concludes.

2 Microfounded forecasting under a general risk function

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. For the moment, we restrict these forecasts to be the result of an opinion poll on the target variable. 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.

We label individual forecasts of y_t, computed using information sets lagged h periods, by $f_{i,t}^h$, $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. Therefore, $f_{i,t}^h$ are h-step-ahead forecasts of y_t, formed at period $t - h$, and N is the number of respondents of this survey of expectations regarding y_t. In this section, we show that, in a variety of interesting cases, optimal forecasts are related to $\mathbb{E}_{t-h}(y_t)$ – the conditional expectation of y_t computed using information lagged h periods – by an affine function of the form[3]:

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

As is well known, Granger (1969) was a pioneer in the forecasting literature. He considered that forecasters minimize a cost function, and that "cost functions that arise in practice in economics and management situations are not likely to be quadratic in form, and frequently will be non-symmetric." If the cost function is symmetric, and additional regularity conditions hold for the density of y_t, then $\mathbb{E}_{t-h}(y_t)$ is obtained as the optimal forecast. In some special cases, optimal forecasts require a bias-correction

[3]Indeed, (1) is an encompassing model. Some results we derive below represent restrictions on this general formula.

term as in $\mathbb{E}_{t-h}(y_t) + k_i^h$.

The subsequent literature on forecast optimization, e.g., Christoffersen and Diebold (1997), Elliott and Timmermann (2004), Patton and Timmermann (2007), and Elliott, Komunjer and Timmermann (2008), have shown the inappropriateness of using the conditional mean under a more general setup, which includes the use of an asymmetric loss function and even an unknown loss function.

In what follows, we consider a setup which has two layers of decisions to be made. In the first layer, individuals (survey respondents) form their optimal point forecasts of a random variable y_t by using a specific *loss function*, unknown to the econometrician, with no knowledge of the DGP of the target variable. These optimal forecasts $f_{i,t}^h$ will be available as survey results, where the number of respondents is either limited ($N < \infty$) or potentially large, $N \to \infty$, depending on the context. Moreover, these surveys can be periodically taken on a large enough number of different occasions, i.e., $T \to \infty$. The setup where $N < \infty$ and $T \to \infty$ describes reasonably well current available surveys of expectations: The Survey of Professional Forecasters, The Michigan Survey, The Livingston Survey, The Bloomberg Survey of Economists, etc. The setup where $N, T \to \infty$ describes reasonably well surveys of the future, covered by the *big data* literature.

In the second layer of decisions, an econometrician will be the final user of this large number of forecasts. We assume that she/he operates under an MSE *risk function*, but this assumption can be modified if need be. However, we believe it covers reasonably well some interesting cases, which are of practical importance, e.g., the government, a central bank, a large risk-neutral firm, etc. Hence, her/his optimal forecast in this second layer of decision making is $\mathbb{E}_{t-h}(y_t)$.

The challenge here is to uncover $\mathbb{E}_{t-h}(y_t)$ – the optimal forecast of the second layer of decision making – from a potentially large set of survey responses, where the econometrician has no knowledge of the loss functions used in the first layer and has also no knowledge of the DGP of the target variable. To make the problem tractable we assume that the unknown DGP of the target variable can be reasonably well approximated by *the location-scale model* with unknown parameters, that could be consistently estimated using standard quantile-regression techniques. As is well known, the location-scale model can approximate several densities used in practice, such as the Normal, Elliptical, Cauchy, Uniform, Logistic, Laplace, Student's t, Generalized extreme value distribution, etc. Moreover, it accommodates most common volatility processes with time-varying variance, e.g., ARCH and stochastic volatility.

The core idea

Before formalizing the assumptions needed to derive our main results, we believe the reader will benefit from a basic explanation of what this paper is about. Consider the setup in Patton and Timmermann (2007), where each individual i chooses the optimal point forecast $f_{i,t}^h$ so that the conditional expected loss function (L^i) is minimized:

$$f_{i,t}^h = \arg\min_f \; \mathbb{E}\left[L^i(y_t; f)\,|\mathcal{F}_{t-h}\right], \tag{2}$$

where $f \in \mathbb{R}$ are all possible choices of the forecaster i, the conditioning information set for all forecasters is the same (Homogeneous Beliefs) and is based on public information, there is an unknown risk function used by forecaster i (Heterogeneity), and an unknown conditional DGP of y_t used by the agent to compute $\mathbb{E}\left[L^i(y_t; f)\,|\mathcal{F}_{t-h}\right]$ and therefore to forecast y_t.

We assume that the optimizing agent employs the location-scale model with one covariate alone, where the conditional quantile function of y_t, $Q_{y_t}(\tau\,|\mathcal{F}_{t-h})$, reads as:

$$Q_{y_t}(\tau\,|\mathcal{F}_{t-h}) = \alpha_0(\tau) + \alpha_1(\tau)\,X_{t-h}, \tag{3}$$

where X_{t-h} is the covariate used in the location-scale model, $\tau \in [0,1]$ is a quantile of the conditional distribution of y_t, and $\alpha_0(\tau)$ and $\alpha_1(\tau)$ are parameters that vary across quantiles.

An important result from Patton and Timmermann (2007) regarding an optimal point forecast $f_{i,t}^h$ is that it is associated with a given quantile level of $Q_{y_t}(\tau\,|\mathcal{F}_{t-h})$, labelled here as τ^*, as follows:

$$f_{i,t}^h = Q_{y_t}(\tau^*\,|\mathcal{F}_{t-h}) = F_{t|t-h}^{-1}(\tau^*)\,. \tag{4}$$

In words, the point forecast $f_{i,t}^h$ is a conditional quantile of y_t, where $F_{t|t-h}(\cdot)$ is the conditional cumulative density function of y_t. The key here is that the point forecast lies in the domain of the conditional density, so that the conditional probability of $y_t \leq f_{i,t}^h$ could be computed and $F_{t|t-h}(\cdot)$ inverted. Given the location-scale model in equation (3), we can apply (4) to obtain:

$$f_{i,t}^h = Q_{y_t}(\tau^*\,|\mathcal{F}_{t-h}) = \alpha_0(\tau^*) + \alpha_1(\tau^*)\,X_{t-h}. \tag{5}$$

Notice that this establishes an affine relationship between the covariate in the location-scale model and the optimal forecast. However, we can go one step further to relate the latter with the conditional mean of y_t, $\mathbb{E}(y_t\,|\mathcal{F}_{t-h}) \equiv \mathbb{E}_{t-h}(y_t)$. This could be accomplished by using another important result (Koenker, 2005), which relates the conditional quantile function with the conditional mean $\mathbb{E}_{t-h}(y_t)$:

$$\begin{aligned}\int_0^1 Q_{y_t}\left(\tau\,|\mathcal{F}_{t-h}\right)d\tau &= \mathbb{E}_{t-h}(y_t), \text{ or,}\\ \mathbb{E}_{t-h}(y_t) &= \int_0^1 \left[\alpha_0\left(\tau\right)+\alpha_1\left(\tau\right)X_{t-h}\right]d\tau\\ &= \int_0^1 \alpha_0\left(\tau\right)d\tau + X_{t-h}\int_0^1 \alpha_1\left(\tau\right)d\tau, \text{ or,}\\ &= \overline{\alpha_0}+\overline{\alpha_1}X_{t-h}, \qquad (6)\end{aligned}$$

where $\overline{\alpha_0}=\int_0^1 \alpha_0\left(\tau\right)d\tau$ and $\overline{\alpha_1}=\int_0^1 \alpha_1\left(\tau\right)d\tau$.

Combining (5) and (6), we arrive at:

$$f_{i,t}^h = \left(\alpha_0\left(\tau^*\right)-\frac{\alpha_1\left(\tau^*\right)\overline{\alpha_0}}{\overline{\alpha_1}}\right)+\frac{\alpha_1\left(\tau^*\right)}{\overline{\alpha_1}}\mathbb{E}_{t-h}(y_t), \qquad (7)$$

which establishes an affine relationship between the optimal point forecast and the conditional expectation $\mathbb{E}_{t-h}(y_t)$. Here, there is a factor model with a sole unobservable factor $\mathbb{E}_{t-h}(y_t)$, which obeys an affine structure. As we show below, this allows filtering out $\mathbb{E}_{t-h}(y_t)$ from a survey of forecasts, since the forecasts $f_{i,t}^h$ are observable.

We can use (7) to check well known results in the literature. Suppose the loss function being used is quadratic. Then, the optimal forecast is the conditional mean. From the location-scale model we obtain $\alpha_1\left(\tau^*\right)=\overline{\alpha_1}$ and $\alpha_0\left(\tau^*\right)=\overline{\alpha_0}$. Therefore, (7) yields:

$$f_{i,t}^h = \mathbb{E}_{t-h}(y_t).$$

Additionally, if the conditional density function of y_t is symmetric, then, $\alpha_1\left(\tau^*\right)=\alpha_1\left(1/2\right)=\overline{\alpha_1}$, $\alpha_0\left(\tau^*\right)=\alpha_0\left(1/2\right)=\overline{\alpha_0}$. Using these results in (7) gives:

$$f_{i,t}^h = \mathbb{E}_{t-h}(y_t) = \text{Median}_{t-h}\left(y_t\right).$$

It is worth noting that, in this context, heterogeneity on the coefficients of (7) could be obtained if every agent uses a different loss function. Another potential source of heterogeneity is the use of a different model for the unknown DGP of y_t across agents. Notice that we have imposed that all agents use the same location-scale model here, so, using a different loss function is the only source of heterogeneity here.

Main assumptions under public information

Assume that individuals $i=1,2,\cdots,N$ forecast y_t conditional on $\mathcal{F}_{t-h}$, where $\mathcal{F}_{t-h}$ is the conditioning set used by all the agents. Optimal forecasts $\widetilde{f}_{i,t}^h$ are obtained by minimizing the respective expected loss function L^i, e.g., Granger and Newbold

(1986).

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

The optimal (point) individual forecasts of y_t are obtained as follows:

$$\widetilde{f}^h_{i,t} \equiv \arg\min_f \ \mathbb{E}\left(L^i(y_t; f) \mid \mathcal{F}_{t-h}\right) \tag{8}$$

where $f \in \mathbb{R}$ are all possible choices of the i-th forecaster and $\mathbb{E}(. \mid \mathcal{F}_{t-h})$, or $\mathbb{E}_{t-h}(.)$, denotes the conditional expectation given $\mathcal{F}_{t-h}$.

An important issue regarding the information set used by each agent is that they all use the same information, which should be regarded as *public information*. Hence, so far, there is no *private information* in our context, and different forecasts arise because agents use a different loss function. Later on, we generalize this setup by also considering private information and heterogeneous beliefs. Also, we treat public information as free, since the minimization problem in (8) has no constraints regarding its cost.

Another issue regarding the choice in (8) is that it is made with no knowledge of the data-generating process of the target variable y_t. Therefore, individual i must use tools to approximate the density function of y_t to be able to solve the problem. Below, we make standard assumptions on how the individual will approximate the density function. We also assume that individual i alone has knowledge on the loss function $L^i(y_t; f)$ used to solve (8). However, standard assumptions on the shape of the loss function will be made as well. On that regard, a natural assumption is that if one forecasts without error, then no forecast loss arises, but, if there is an error, the larger it is, the greater will be the loss:

Assumption A2 (Shape of the loss function) 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

[4] This is the same Assumption L1 of Patton and Timmermann (2007). According to them, 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.

class of loss functions. A simple way to consider an asymmetric function, and account for the "degree of asymmetry", is given by the following assumption:

Assumption A3 (Asymmetry of the loss function) 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.

Assumptions A2 and A3 are standard in the literature (e.g. Granger and Newbold, 1986; Patton and Timmermann, 2007). Note that A3 includes the symmetric case when $\beta_1 = \beta_2$. It 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.

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

Assumption A4 is a mild technical condition on the DGP of y_t.

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.[5]

Assumption 5 deals with how an individual i will approximate the conditional DGP of y_t. Notice that no parametric structure is placed on $F_{\eta,h}(\cdot)$ and the covariates affect both the location and the scale of the conditional distribution of y_t. Moreover, A5 implies that: (i) $Q_{y_t}(\tau\,|\mathcal{F}_{t-h}) = \alpha_0(\tau) + \alpha_1(\tau)x_{t,t-h}$ 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 $Q_{y_t}(.)$ is the conditional quantile function of y_t, $[\alpha_0(\tau); \alpha_1(\tau)]$ depends on $(\delta;\gamma)$, L^i and $F_{\eta,h}(0,1)$; and $\overline{\alpha_j} \equiv \int_0^1 \alpha_j(\tau)d\tau$

[5] For ease of notation, we assume that $X_{t,t-h}' = (1, x_{t,t-h})$ is a 2×1 vector, $\delta = (\delta_0, \delta_1)'$, and $\gamma = (\gamma_0, \gamma_1)'$. See the Online Technical Appendix (section 5) for some possible ways to consider the multi-covariates case.

for $j = \{0; 1\}$. The previous expressions for both the conditional quantile and the conditional expectation of y_t (under A5) will be exploited next to deliver a linear connection between the optimal forecast and the conditional mean. First, we show this when we assume $\alpha_0(\tau)$ and $\alpha_1(\tau)$ known. Then, we show how to deal with estimation of $\alpha_0(\tau)$ and $\alpha_1(\tau)$ and its consequences.

Proposition 1 *(Location-scale model) If A1-A5 hold, then: (i) the optimal forecast is an affine 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 $\beta_i^h = 1$, for all i, i.e., $\widetilde{f}_{i,t}^h = k_i^h + \mathbb{E}_{t-h}(y_t)$.*

In practice, only the survey respondent knows her/his own loss function, but not the econometrician. Moreover, it is not realistic to assume that either the respondent or the econometrician know the DGP of y_t. Therefore, an optimal forecast of the form:

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

obtained above is not feasible, since k_i^h and β_i^h are unknown parameters and $\mathbb{E}_{t-h}(y_t)$ is a latent variable.

A natural way for survey respondents to estimate (9) is by using quantile-regression techniques, pioneered by Koenker and Bassett (1978), with further developments discussed in Koenker (2005). Indeed, under standard assumptions about the data, which are likely to hold in our context, it is possible for the survey respondent to estimate consistently k_i^h, β_i^h, and $\mathbb{E}_{t-h}(y_t)$. If we denote by $\widehat{k_i^h}$, $\widehat{\beta_i^h}$, and $\widehat{\mathbb{E}}_{t-h}(y_t)$ the respective consistent estimates of k_i^h, β_i^h, and $\mathbb{E}_{t-h}(y_t)$, the optimal (feasible) survey responses are:

$$f_{i,t}^h = \widehat{k_i^h} + \widehat{\beta_i^h} \widehat{\mathbb{E}}_{t-h}(y_t). \tag{10}$$

Notice that we have made the distinction between the optimal forecast $\widetilde{f}_{i,t}^h$ and its optimal feasible counterpart $f_{i,t}^h$ which requires the estimation of k_i^h, β_i^h, and $\mathbb{E}_{t-h}(y_t)$, i.e., equation (9) versus (10). We now use a technical assumption that guarantees the existence of quantile-regression consistent estimates used in (10).

Assumption A6 Define $[\widehat{k_i^h}; \widehat{\beta_i^h}] = [\widehat{\alpha}_0(\tau_i) - \frac{\widehat{\overline{\alpha_0}}}{\widehat{\overline{\alpha_1}}} \widehat{\alpha}_1(\tau_i); \frac{\widehat{\alpha}_1(\tau_i)}{\widehat{\overline{\alpha_1}}}]$, where $[\widehat{\alpha}_0(\tau_i); \widehat{\alpha}_1(\tau_i)]$ are the resulting estimates (intercept and slope) of a standard linear quantile regression of y_t onto $[1; x_{t,t-h}]$ at quantile level τ_i. In addition, let the average coefficients $\widehat{\overline{\alpha_j}} = \sum_{k=1}^{K} \widehat{\alpha}_j(\tau_k) \Delta\tau_k$, for $j = \{0; 1\}$, be computed as Riemann sums over a grid of K equidistant quantile levels $\tau_k \in [\tau_1, \tau_2, ..., \tau_K]$, such that $\tau_k = \frac{k}{K+1}$ and $\Delta\tau_k = \frac{1}{K+1}$ for $k = [1, ..., K]$. Also assume that regularity conditions A1-A2 of Koenker (2005, p.120) on $x_{t,t-h}$ are met, and that $\alpha(\tau)$ is continuous and Riemann-integrable on $[0, 1]$. Finally, define $\widehat{\mathbb{E}}_{t-h}(y_t) \equiv \widehat{\overline{\alpha_0}} + \widehat{\overline{\alpha_1}} x_{t,t-h}$.

Proposition 2 *If A1-A6 hold, then, the optimal (feasible) forecast of* y_t *conditional on* $\mathcal{F}_{t-h}$ *is of the form:* $f^h_{i,t} = k^h_i + \beta^h_i \mathbb{E}_{t-h}(y_t) + \varepsilon^h_{i,t}$*, where* $\varepsilon^h_{i,t}$ *accounts for finite sample parameter uncertainty, and* $[\widehat{k^h_i}; \widehat{\beta^h_i}]$ *are consistent estimates of* $[k^h_i; \beta^h_i]$*.*

Notice that, adding and subtracting $k^h_i + \beta^h_i \mathbb{E}_{t-h}(y_t)$ to (10), leads to:

$$f^h_{i,t} = k^h_i + \beta^h_i \mathbb{E}_{t-h}(y_t) + \varepsilon^h_{i,t}, \tag{11}$$

where it becomes clear that $\varepsilon^h_{i,t} = \left[\widehat{k^h_i} - k^h_i\right] + \left[\widehat{\beta^h_i}\widehat{\mathbb{E}}_{t-h}(y_t) - \beta^h_i \mathbb{E}_{t-h}(y_t)\right]$ reflects the small-sample error in approximating the unknown DGP[6]. There are several special cases of interest where the optimizing agent knows the DGP. To save space, we discuss them on the Online Technical Appendix to this paper.

As is clear from above, quantile-regression techniques allowed the estimation of $\widehat{\mathbb{E}}_{t-h}(y_t)$, $\widehat{k^h_i}$ and $\widehat{\beta^h_i}$, but this is only available for the individual forecaster in the first layer of decision making, since we assumed that only the survey respondent knows her/his own loss function, but not the econometrician in the second layer. Indeed, the survey respondent will deliver a bias and error ridden conditional expectation to the final user of the survey in the second layer – the econometrician, i.e., equation (11). So, from the point of view of the second layer, we are facing a typical signal-extraction problem, which we solve exploiting the fact that all forecasts are functions of the common factor $\mathbb{E}_{t-h}(y_t)$, leading naturally to errors that do not depend on $\mathcal{F}_{t-h}$. This can be the basis of a generalized-method-of-moment estimate of $\mathbb{E}_{t-h}(y_t)$ as we discuss next.

Identification and GMM estimation under public information alone

Consider an econometrician who only observes a survey of individual forecasts $f^h_{i,t}$ (all optimal and feasible, in principle) and the target variable y_t, but has no information at all about the DGP and the individual loss functions previously used in the first layer of decision making. Because the class of location-scale models is relatively broad, we assume that the econometrician will employ this class of models to approximate the unknown DGP[7]. From our previous results, it is natural for the econometrician to use

[6] An optimal forecast "approximation error" $\varepsilon^h_{i,t} \equiv f^h_{i,t} - \widetilde{f}^h_{i,t}$ arises in (11) when the DGP is unknown. In the Online Technical Appendix (section 4), we provide a full account of these error terms using as examples a few parametric DGPs.

[7] The use of the location-scale class can be generalized, in principle, but we leave this for future research.

the derived factor model in Proposition 2:

$$f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h. \tag{12}$$

Equation (12) is a three-dimensional panel, with $t = 1, ..., T$, $i = 1, ..., N$, and $h = 1, ..., H$. Although the horizon can in principle increase without bound, when forecasting one usually keeps H small. Recall that, in a stationary-ergodic world, as H increases, conditional forecasts converge to their unconditional counterparts, making the case for large H unappealing.

Our framework entails two interesting cases. The first is $T \to \infty$, with small N and H, which applies to some long standing surveys with an almost constant number of respondents throughout time, e.g., the *Livingston Survey* of the Philadelphia FED, available since 1959 on a biannual basis, the *Survey of Professional Forecasters*, available since 1968 on a quarterly basis, or the *Wall Street Journal* Survey of Forecasters, available since 2003 on a monthly basis. The second case is $T \to \infty$, $N \to \infty$, and small H, which applies to large surveys taken across a long period of time. At the time of writing, we see no surveys of expectations that could fit this setup. But, this does not mean that we should ignore it. In a world where there is an increasing availability of reliable data provided electronically, it is interesting to examine how one could efficiently use this wealth of information for forecasting. Of course, there is already a large literature on *big data* in statistics, but it could benefit from economic interpretation offered by the (structural) factor model discussed above. In the future, surveys of expectations would include the abundant information on thousands or millions of respondents on an almost continuous-time basis. This makes the *big data* setup where $T \to \infty$, $N \to \infty$, and small H, of interest to be applied in the *Surveys of the Future*, being the reason why we discuss them here.

There is one hurdle to overcome when N and T are large – the curse of dimensionality when $N \to \infty$. Even when H is small, the system of equations (12) has too many parameters that grow without bound, which pose a problem for the identification of $\mathbb{E}_{t-h}(y_t)$. However, we must stress that we do not need the identification of all k_i^h and β_i^h, $i = 1, ..., N$, to be able to identify $\mathbb{E}_{t-h}(y_t)$. Indeed, under suitable conditions, we only need to know their respective means. Averaging (12) across i, and assuming that

$\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^{h} \xrightarrow{p} 0$, allows identifying $\mathbb{E}_{t-h}(y_t)$ as a function of three means alone:

$$\mathbb{E}_{t-h}(y_t) = \operatorname*{plim}_{N\to\infty}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N} k_i^h}{\frac{1}{N}\sum_{i=1}^{N}\beta_i^h}\right), \tag{13}$$

as long as all the terms in parenthesis in (13) converge in probability. The only problem here is that $\overline{k^h} = \frac{1}{N}\sum_{i=1}^{N} k_i^h$ and $\overline{\beta^h} = \frac{1}{N}\sum_{i=1}^{N}\beta_i^h$ are not known. Next, we discuss their consistent estimation.

Our proposed approach to identify and estimate $\mathbb{E}_{t-h}(y_t)$ is to employ the generalized method of moments (GMM), relying on T asymptotics. However, the fact that $\mathbb{E}_{t-h}(y_t)$ is a latent variable is a serious drawback, since the moments used in GMM estimation must be a function of observables and parameters alone. Following Issler and Lima (2009), one can always decompose the series y_t into $\mathbb{E}_{t-h}(y_t)$ and an unforecastable martingale-difference component η_t^h, such that $\mathbb{E}_{t-h}(\eta_t^h) = 0$:

$$y_t = \mathbb{E}_{t-h}(y_t) - \eta_t^h. \tag{14}$$

Thus, combining (12) and (14) leads to:

$$\begin{aligned} f_{i,t}^h &= k_i^h + \beta_i^h(y_t + \eta_t^h) + \varepsilon_{i,t}^h &\quad (15)\\ &= k_i^h + \beta_i^h y_t + v_{i,t}^h, &\quad (16)\end{aligned}$$

where $v_{i,t}^h \equiv \beta_i^h\eta_t^h + \varepsilon_{i,t}^h$ is a composite error term. Notice that, by construction, $\mathbb{E}\left(\eta_t^h \mid \mathcal{F}_{t-h}\right) = 0$, so, $\mathbb{E}\left(v_{i,t}^h \mid \mathcal{F}_{t-h}\right) = 0$ if one assumes that $\mathbb{E}(\varepsilon_{i,t}^h \mid \mathcal{F}_{t-h}) = 0$. In the context where $T, N \to \infty$, it is also reasonable to expect that the approximation error $\varepsilon_{i,t}^h$ vanishes, since it reflects the small-sample error in approximating the unknown DGP of y_t.

Starting with (16) and $\mathbb{E}\left(v_{i,t}^h \mid \mathcal{F}_{t-h}\right) = 0$, and using the law of iterated expectations and valid observable instruments z_{t-s}, where $z_{t-s} \in \mathcal{F}_{t-h}$, $s \geq h$, we obtain:

$$\mathbb{E}\left[\left(f_{i,t}^h - k_i^h - \beta_i^h y_t\right) \otimes z_{t-s}\right] = 0, \tag{17}$$

which is valid for all $i = 1, ..., N$, $t = 1, ..., T$, and $h = 1, ..., H$. The system (17) has $2NH$ parameters and (at least) $2NH$ moment conditions, provided that $dim(z_{t-s}) \geqslant 2$, which is critical for identification. Despite that, one problem remains: as $N \to \infty$, the amount of parameters in (17) diverges, which goes against consistency. Notice, however, that $T \to \infty$ poses no such problem.

In a somewhat similar framework, Driscoll and Kraay (1998) discuss the *curse of dimensionality* for panel regressions estimated by GMM. Their focus is on the conditions under which spatial dependence implies a controlled environment for time-series dependence. It relies on T-asymptotics and applies for the case where N is fixed and the case where $N \to \infty$. Notice that these are exactly the two interesting cases alluded above. We first discuss $T \to \infty$, $N \to \infty$, with small H. Driscoll and Kraay's solution to the curse of dimensionality was to take cross-sectional averages of terms such as $\left(f_{i,t}^h - k_i^h - \beta_i^h y_t\right) \otimes z_{t-s}$. In their context, the parameters to be estimated by GMM did not depend on i, although the data did. Here, both parameters and the data depend on i. Despite that, one can still use cross-sectional averages to reduce dimensionality[8], leading to:

$$\mathbb{E}\left[\left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \otimes z_{t-s}\right] = 0, \tag{18}$$

$t = 1, ..., T$, and $h = 1, ..., H$, where $\overline{f_{\cdot,t}^h} = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$, $\overline{k^h} = \frac{1}{N}\sum_{i=1}^{N} k_i^h$ and $\overline{\beta^h} = \frac{1}{N}\sum_{i=1}^{N} \beta_i^h$, represent cross-sectional averages for each h.

As argued above, since we need not know the individual coefficients k_i^h and β_i^h, but only their means to be able to identify and estimate $\mathbb{E}_{t-h}(y_t)$ from a survey of forecasts, and since $\mathbb{E}_{t-h}(y_t)$ does not vary across i, averaging across i as in (18) is an interesting strategy to recover $\mathbb{E}_{t-h}(y_t)$. As long as these cross-sectional averages converge, GMM using time-series restrictions delivers consistent estimates of the respective parameter means, under standard conditions for the consistency of GMM estimates. Once one deals successfully with consistency, one can start worrying about efficiency in a GMM context.

One way to exploit all possible moment conditions implicit in (18) is to stack all the restrictions across h (finite) as:

$$\mathbb{E}\left[\begin{pmatrix} \overline{f_{\cdot,t}^1} - \overline{k^1} - \overline{\beta^1} y_t \\ \overline{f_{\cdot,t}^2} - \overline{k^2} - \overline{\beta^2} y_t \\ \vdots \\ \overline{f_{\cdot,t}^H} - \overline{k^H} - \overline{\beta^H} y_t \end{pmatrix} \otimes z_{t-s}\right] = 0, \tag{19}$$

where our problem collapses to one where we have $H \times dim(z_{t-s})$ restrictions and $2H$ parameters to estimate. As before, over-identification requires that $dim(z_{t-s}) > 2$.

[8] Provided that z_{t-s} does not depend on i, which is implied by our notation.

Given a choice of H, GMM estimation of (19) is efficient. A less efficient alternative to estimate the whole stacked system (19) is to estimate separately (18) for every horizon h, which could also be attempted for computational reasons.

In the context of $N, T \to \infty$, with H small, we now discuss the case where we first let $N \to \infty$ and then let $T \to \infty$, using the sequential asymptotic framework of Phillips and Moon (1999). Under suitable conditions, the cross-sectional averages in (18) and (19) would converge in probability to a unique limit as $N \to \infty$, i.e., $\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} \beta_i^h = \beta^h$, $\beta^h \neq 0$, $\left|\beta^h\right| < \infty$, $\text{plim} \frac{1}{N}\sum_{i=1}^{N} k_i^h = k^h$, $\left|k^h\right| < \infty$, and $\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h = f_{\cdot,t}^h$, $\left|f_{\cdot,t}^h\right| < \infty$, for all $t = 1, 2, \ldots, T$. Then, a key condition for T-consistent estimation is that (18) and (19), evaluated at these limits are a unique solution for each of them respectively. Take now moments (18), considering that $N \to \infty$:

$$\mathbb{E}\left[\left(f_{\cdot,t}^h - k^h - \beta^h y_t\right) \otimes z_{t-s}\right] = 0, \text{ or,} \tag{20}$$

$$\mathbb{E}\left[\left(\beta^h \eta_t^h\right) \otimes z_{t-s}\right] = 0, \tag{21}$$

under $\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t}^h \xrightarrow{p} 0$, as $N \to \infty$, $t = 1, ..., T$, and $h = 1, ..., H$. Note that (21) is attained by construction, since η_t^h is a martingale-difference and must be orthogonal to all series dated $t-h$ or before. In this context, GMM provides a consistent estimate for parameter means as $T \to \infty$. To prove it, first define $\theta^h = [k^h; \beta^h]'$ and consider the following assumptions:

Assumption A7 Let $\boldsymbol{\varepsilon}_t^h = \left(\varepsilon_{1,t}^h, \varepsilon_{2,t}^h, \ldots, \varepsilon_{N,t}^h\right)'$ 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 $\mathbb{E}\left(\varepsilon_{i,t}^h\right) = 0$ for all i and t, given h. Furthermore, we assume that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N^2} \sum_{i=1}^{N}\sum_{j=1}^{N} \left|\mathbb{E}\left(\varepsilon_{i,t}^h \varepsilon_{j,s}^h\right)\right| = 0, \tag{22}$$

for all t and s, given h.

Assumption A8 We assume that $\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} \beta_i^h = \beta^h \neq 0$, $\left|\beta^h\right| < \infty$, $\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} k_i^h = k^h$, $\left|k^h\right| < \infty$, and $\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h = f_{\cdot,t}^h$, $\left|f_{\cdot,t}^h\right| < \infty$, for all $t = 1, 2, \ldots, T$.

Assumption A9 We assume that the identification conditions for GMM estimation are met and that there is a unique set of values $\theta_0^h = [k_0^h; \beta_0^h]'$, $h = 1, 2, \ldots, H$, that solve either (19) or (18) for each h separately[9]. We further assume that the additional regularity conditions used by Hansen (1982) in proving T-consistency of GMM estimates $\widehat{\theta^h} = [\widehat{k^h}; \widehat{\beta^h}]'$ are met as well.

Assumption A7 guarantees that the errors $\varepsilon_{i,t}^h$ can be diversified away, and that cross-sectional dependence is not a problem. It is required in a GMM context in order to ensure that $\mathbb{E}_{t-h}\left(\underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N} v_{i,t}^h\right) = 0$. This serves as a basis to obtain either (18) or (19), when $N \to \infty$. Notice that, although this is an assumption here, it can be tested using standard over-identifying restrictions tests. Assumption A8 just requires finite convergence of different cross-sectional averages, which bounds the degree of cross-sectional and time-series dependence due to spatial dependence. They are expected to hold on a stationary-ergodic context. Assumption A9 deals with GMM identification and is standard in the literature. We can now state an important result, which allows for consistent estimation of $\mathbb{E}_{t-h}(y_t)$.

Proposition 3 *If A1-A9 hold, then, the feasible Extended BCAF (Bias Corrected Average Forecast)* $\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{k^h}}{\widehat{\beta^h}}$, *based on T-consistent GMM estimates* $\widehat{\theta^h} = \left[\widehat{k^h}; \widehat{\beta^h}\right]'$, *obeys* $\underset{(N,T\to\infty)_{\text{seq}}}{\text{plim}} \left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{k^h}}{\widehat{\beta^h}}\right) = \mathbb{E}_{t-h}(y_t)$, *where $(N,T \to \infty)_{\text{seq}}$ denotes the sequential asymptotic approach proposed by Phillips and Moon (1999), when we let first $N \to \infty$, and then let $T \to \infty$.*

We now turn into the complicated case where in the sequential asymptotics we let first $T \to \infty$, and then $N \to \infty$, or, that we let $T \to \infty$ with N fixed. The critical issue of letting $T \to \infty$ first is that, when we take cross-sectional averages as in (18) or (19), but N is not large, we have to guarantee that for all possible ensemble averages $\mathbb{E}_{t-h}(y_t)$ is identified. When we assume the opposite sequential order, we have only to worry about identification of a single limit case.

An added complication is that $\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t}^h$ does not vanish in probability. Despite that, A7 imposes that $\mathbb{E}\left(\varepsilon_{i,t}^h\right) = 0$ for all i and t, given h. This is not sufficient to guarantee that moments (18) and (19) can be cast in terms of η_t^h alone. However, employing the more restrictive assumption that $\mathbb{E}(\varepsilon_{i,t}^h \mid \mathcal{F}_{t-h}) = 0$ is sufficient to validate (18) and (19). Thus, we employ the following assumptions outlined next:

Assumption A10 Let $\boldsymbol{\varepsilon}_t^h = \left(\varepsilon_{1,t}^h, \varepsilon_{2,t}^h, \ldots, \varepsilon_{N,t}^h\right)'$ be a $N \times 1$ vector stacking the errors $\varepsilon_{i,t}^h$ associated with all possible forecasts. Assume that the vector process $\left\{\boldsymbol{\varepsilon}_t^h\right\}$

[9] Since the restrictions are linear, the θ_0^h, $h = 1, 2, \ldots, H$, are identified up to nonzero scalar multiplication.

is covariance-stationary and ergodic for the first and second moments, uniformly on N, and that $\mathbb{E}(\varepsilon_t^h \mid \mathcal{F}_{t-h}) = 0$ for all t, given h.

Assumption A11 Define $\frac{1}{N}\sum_{i=1}^{N} \beta_i^h = \overline{\beta^h}$ and $\frac{1}{N}\sum_{i=1}^{N} k_i^h = \overline{k^h}$. We assume that, for all N, the identification conditions for GMM estimation are met and that there is a unique set of values $\overline{\theta_0^h} = [\overline{k_0^h}; \overline{\beta_0^h}]'$, $h = 1, 2, \ldots, H$, that solves either (19) or (18) for each h separately. We further assume that the additional regularity conditions used by Hansen (1982) in proving consistency of GMM are met as well.

We now turn to the analogous estimation result when we reverse the order of sequential asymptotics: first we let $T \to \infty$ with N fixed. Later, we let $T \to \infty$ and then $N \to \infty$. The latter case can represent a *big data* environment, while the former clearly applies to current surveys of expectations.

Proposition 4 *If A1-A6 and A10-A11 hold, then, the feasible Extended BCAF (Bias Corrected Average Forecast)* $\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}$*, based on T-consistent GMM estimates* $\widehat{\overline{\theta^h}} = \left[\widehat{\overline{k^h}}; \widehat{\overline{\beta^h}}\right]'$*, obeys* $\underset{T\to\infty}{plim}\left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right) = \mathbb{E}_{t-h}(y_t)$*, where we let $T \to \infty$, with N fixed. The convergence to $\mathbb{E}_{t-h}(y_t)$ also happens when we let first $T \to \infty$ and then let $N \to \infty$, that is,* $\underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right) = \mathbb{E}_{t-h}(y_t)$*, where $(T, N \to \infty)_{\text{seq}}$ denotes the sequential asymptotic approach proposed by Phillips and Moon (1999).*

Notice that assumptions A10 and A11 suffice to provide T-consistent GMM estimates for all bounded N. Indeed, fixed N is a special case of $(T, N \to \infty)_{\text{seq}}$, where we do not let N diverge after $T \to \infty$, the context being identical otherwise. For the sake of completeness, it would be interesting to cover the case of joint convergence, where we can establish a link between sequential convergence and joint convergence following Phillips and Moon (1999). However, to save space in the main text, this is done in the Online Technical Appendix of this paper.

Identification and GMM estimation under public and private information

According to equation (1), each forecast is an affine function of the conditional expectation, and the informational content of each forecast over time is identical. In other words, the loss functions of the individual forecasters are assumed to be heterogeneous, but the estimation methods for the implementation of their optimal forecast are based (so far) on the same information set. In practice, however, each survey respondent could use different covariates to compute their optimal forecast, besides adding expert judgment to their econometric models.

To address this relevant issue, we now add an extra source of heterogeneity to the basic setup, by allowing individuals to have different information sets $\mathcal{F}_{i,t-h}$ when forming their optimal (feasible) forecasts. It is straightforward to show that, following our steps in the previous section, we obtain:

$$f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{i,t-h}(y_t) + \varepsilon_{i,t}^h, \tag{23}$$

where now agents have also a heterogeneous information set $\mathbb{E}(y_t \mid \mathcal{F}_{i,t-h}) \equiv \mathbb{E}_{i,t-h}(y_t)$.

In what follows, the new information set $\mathcal{F}_{i,t-h}$ is partitioned into three mutually orthogonal (and non-empty) components: $\mathcal{F}_{t-h}$, $\mathcal{F}_{t-h}^{PC}$, and $\mathcal{F}_{i,t-h}^{PI}$, for all $i = 1, ..., N$, where $\mathcal{F}_{i,t-h} = \mathcal{F}_{t-h} \cup \mathcal{F}_{t-h}^{PC} \cup \mathcal{F}_{i,t-h}^{PI}$. Again, the public information set is $\mathcal{F}_{t-h}$, i.e., the information available to all forecasters. Private information is in turn decomposed into the private-common information set $\mathcal{F}_{t-h}^{PC}$ (e.g., paid-up news/macro reports, unobserved market consensus, etc.) and the private-idiosyncratic information set $\mathcal{F}_{i,t-h}^{PI}$ (e.g., idiosyncratic private information, individual expert judgement, etc.).

One important result widely used in the *Nowcasting* literature, e.g., Bánbura, Gianone, and Reichlin (2011), is the orthogonal-component decomposition:

$$\mathbb{E}_{i,t-h}(y_t) \equiv \mathbb{E}(y_t \mid \mathcal{F}_{i,t-h}) = \mathbb{E}(y_t \mid \mathcal{F}_{t-h}) + \mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right) + \mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{PI}\right), \tag{24}$$

where it becomes clear that heterogeneity only appears in the last term of the decomposition, while $\mathbb{E}(y_t \mid \mathcal{F}_{t-h}) + \mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right)$ contains the common part of expectations for all $i = 1, ..., N$. Therefore, if we average (23) using (24), considering that $\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t}^h \xrightarrow{p} 0$ and $\frac{1}{N}\sum_{i=1}^{N} \beta_i^h \mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{PI}\right) \xrightarrow{p} 0$, we can identify the common component $\mathbb{E}(y_t \mid \mathcal{F}_{t-h}) + \mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right)$ as follows:

$$\mathbb{E}(y_t \mid \mathcal{F}_{t-h}) + \mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right) = \operatorname*{plim}_{N\to\infty} \left(\frac{\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i^h}{\frac{1}{N}\sum_{i=1}^{N} \beta_i^h} \right), \tag{25}$$

as long as all the terms in parenthesis in (25) converge in probability. As before, the average parameters in (25) can be consistently estimated by GMM as in Propositions 3 and 4. We present next additional assumptions needed to cover heterogeneous beliefs and our main result in this context.

Assumption A12 Let $\mathcal{F}_{t-h}$, $\mathcal{F}_{t-h}^{PC}$ and $\mathcal{F}_{i,t-h}^{PI}$ be mutually orthogonal and non-empty information sets that form a partition of $\mathcal{F}_{i,t-h}$, where $-\infty < \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{i,t-h}(y_t) \right) < \infty$, for all $t = 1, \ldots, T$ and $h = 1, \ldots, H$, and let $\underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{PI}\right) \right) = 0$, for all $t = 1, \ldots, T$ and $h = 1, \ldots, H$. Define $\eta_{i,t}^h = \mathbb{E}_{i,t-h}(y_t) - y_t$, and let $\eta_t^h = \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \eta_{i,t}^h \right)$ and $\beta^h \eta_t^h = \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \beta_i^h \eta_{i,t}^h \right)$, where $\left|\eta_t^h\right| < \infty$ and $\left|\beta^h \eta_t^h\right| < \infty$, for all $t = 1, \ldots, T$ and $h = 1, \ldots, H$.

Proposition 5 *(Heterogeneous Beliefs) Under assumptions A1-A9 and A12, the optimal (feasible) forecast of* y_t *conditioned on* $\mathcal{F}_{i,t-h}$ *is of the form:* $f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{i,t-h}(y_t) + \varepsilon_{i,t}^h$*, and the feasible Extended BCAF* $\frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{k^h}}{\widehat{\beta^h}}$*, based on* T*-consistent GMM estimates* $\widehat{\theta^h} = \left[\widehat{k^h}; \widehat{\beta^h}\right]'$*, obeys* $\underset{(N,T\to\infty)_{\text{seq}}}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{k^h}}{\widehat{\beta^h}} \right) = \mathbb{E}(y_t \mid \mathcal{F}_{t-h}) + \mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right)$ *where* $(N, T \to \infty)_{\text{seq}}$ *denotes the sequential asymptotic approach of Phillips and Moon (1999), when we let first* $N \to \infty$*, and then let* $T \to \infty$*.*

Proposition 5 allows for heterogeneous information sets and enables us to decompose the survey-based proxy for the conditional expectation (the feasible EBCAF) into public and private terms. Moreover, if all the variables in the public information set $\mathcal{F}_{t-h}$ can be summarized by a factor model, then one can recover the public information term $\widehat{\mathbb{E}}(y_t \mid \mathcal{F}_{t-h})$ as well as the private information term $\widehat{\mathbb{E}}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right) + \widehat{\mathbb{E}}\left(y_t \mid \mathcal{F}_{i,t-h}^{PI}\right)$.[10]

An alternative to GMM estimation to extract the factor is to employ the principal-component analysis proposed in the common factor literature in economics and finance based on the results of Forni et al. (2000), and Stock and Watson (2002b). As is well known, consistent estimation of the common factor requires both N and T to diverge. Therefore, they fit perfectly well the setup of what we have called the *Surveys of the Future*, where $N, T \to \infty$, but not the setup of current surveys of expectations, where $T \to \infty$ but N is fixed. For current surveys GMM estimation has the advantage of only relying on T asymptotics to estimate $\overline{k^h}$ and $\overline{\beta^h}$, regardless of whether $N \to \infty$ or not. In the empirical section we compare the forecasting accuracy of GMM based estimates of $\mathbb{E}_{t-h}(y_t)$ and of $\mathbb{E}(y_t \mid \mathcal{F}_{t-h}) + \mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right)$ with that of principal-component based estimates.

[10]Following our previous steps in Proposition 3 and 4, it is straightforward to consider the case where $T \to \infty$, with N fixed.

Discussion

By applying the tools of the literature on optimal forecasts, and modelling the forecasting decision of an individual from *first principles* (i.e., *microfounded*), we propose a *financial approach* to economic forecasting where we show that surveys of forecasts obey an affine factor structure with a single factor which is the common component of the conditional expectation of the target variable. This holds in a context where individuals only have access to public information or where they also have access to private information with common and idiosyncratic components. The affine structure allows the combination of forecasts to eliminate idiosyncratic forecast errors, supplying superior forecasts under an *MSE risk function*.

Our approach involves two layers of decision making. In the first, individuals choose the best forecast to be posted on the database containing surveys of forecasts. In the second, an *econometrician* uses survey results to forecast the target variable optimally. From the point of view of the econometrician, it is important to estimate the common factor of the survey, which is the optimal forecast under public information and an *MSE risk function*. We show how this can be performed using GMM. We also discuss alternative estimation techniques based primarily on principal-component analysis. In this context, the feasible optimal forecast is a function of the consensus forecast of the survey (a cross-sectional average of survey forecasts) after appropriately filtering out two bias terms. This links the *financial approach* to economic forecasting to the forecast-combination literature. Indeed, if the number of forecasts is large enough, i.e., $N \to \infty$, all idiosyncratic forecast risks are diversified away.

Our results are applicable to two types of surveys with a large enough number of time observations ($T \to \infty$): one of *current surveys of forecasts*, which possess a limited number of respondents ($N < \infty$), and one that we have labelled as the *Surveys of the Future*, where the number of respondents is also large ($N \to \infty$). The latter connects this paper with the literature on *big data*. In standard GMM moment estimation, we circumvent the *curse of dimensionality* that arises from the factor structure (large N) by employing cross-sectional averages. In a big-data context, this allows the use of all the information contained in the survey, while estimating a parsimonious factor model, with only two biases – intercept and slope bias.

Alternative identification strategies could be pursued, especially those that are based on principal-component analysis as proposed by Forni et al. (2000) and Stock and Watson (2002b). Indeed, we also employ these techniques to estimate the optimal forecast in the second layer and compare results with our GMM approach. Another possibility is to employ a within-group method as proposed by Bai (2009). This allows for the joint presence of additive and interactive effects. In our notation, Bai's structure, with public information alone, would be:

$$f_{i,t}^{h} = X_{i,t}\gamma + \mu + k_i^h + \eta_t^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h. \tag{26}$$

One could think that our affine structure could be viewed as a special case of Bai's model, where $X_{i,t} = 0$ and $\mu = 0$. However, identification of Bai's latent factor ($\mathbb{E}_{t-h}(y_t)$ here) requires constraining $\sum_{i=1}^{N} \beta_i^h = 0$ and $\sum_{t=1}^{T} \mathbb{E}_{t-h}(y_t) = 0$, which our method avoids.

One important character of our method is that it extends the previous literature on a *financial approach* to economic forecasting using panel data, e.g., Palm and Zellner (1992), Davies and Lahiri (1995), Davies (2006), Issler and Lima (2009), Lahiri, Peng, and Sheng (2015), and Gaglianone and Lima (2014). For example, our setup encompasses that of Davies and Lahiri (and Davies), reproduced below with our notation:

$$y_t - f_{i,t}^h = -\left(k_i^h + \eta_t^h + \varepsilon_{i,t}^h\right), \tag{27}$$

since it imposes $\beta_i^h = 1$ for all $i = 1, ..., N$ and all $h = 1, ..., H$. Also, it generalizes the results in Issler and Lima, where $\beta_i^h = 1$ for all $i = 1, ..., N$, since they have only one source of bias correction, the intercept.

For a given horizon h, the parameters in $\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t$, i.e., (18), can be estimated using instrumental variables (IV), considering the average forecast $\overline{f_{\cdot,t}^h} = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$ as a dependent variable and the target variable y_t and an intercept as explanatory variables, where the elements in z_{t-s} are all valid instruments. Here, $\overline{f_{\cdot,t}^h}$, called the *consensus forecast*, is the dependent variable, whereas the IV regression recovers the average intercept and slope biases. One can also solve the IV regression for y_t, obtaining an *inverse regression* relating the latter with $\overline{f_{\cdot,t}^h}$, and apply the usual Mincer and Zarnowitz (1969) tests of no bias involving zero intercept and unit slope. Starting from (16), averaging across i, we obtain:

$$\overline{f_{\cdot,t}^h} = \overline{k^h} + \overline{\beta^h} y_t + \overline{v_{\cdot,t}^h}, \text{ or, the inverse regression,} \tag{28}$$

$$y_t = -\frac{\overline{k^h}}{\overline{\beta^h}} + \frac{1}{\overline{\beta^h}}\overline{f_{\cdot,t}^h} - \frac{\overline{v_{\cdot,t}^h}}{\overline{\beta^h}}, \tag{29}$$

where (29) is the usual Mincer-Zarnowitz regression, estimated by least squares, which is used to test for an unbiased consensus: $\overline{k^h} = 0$ and $\overline{\beta^h} = 1$; see the discussion in Gaglianone, Issler and Matos (2017) on identification and estimation. What is striking here is that, in the presence of the factor model (12), one cannot estimate (29) by least squares, since the regressor $\overline{f_{\cdot,t}^h}$ is correlated with the error $\frac{\overline{v_{\cdot,t}^h}}{\overline{\beta^h}}$, due to (28). Therefore, to test for rationality one must use instruments as we proposed above in a GMM setup.

3 Empirical Application

3.1 Data

The Focus Survey of forecasts is a world-class panel database put together by the Central Bank of Brazil (BCB), which collects *daily* information from professional forecasters of commercial banks, asset management firms, consulting firms and non-financial institutions, followed throughout time with a reasonable turnover. As new participants are often added to the survey, and others drop out, the panel of survey forecasts is unbalanced. From a set of 254 registered institutions in the system, there is a smaller *active* group of around 100 institutions that frequently update their nowcasts and forecasts. These are supplied over different forecast horizons and for a large array of macroeconomic time series; see Marques (2013).

Our target variable in this forecasting exercise is Brazilian inflation, as measured by the Broad National Consumer Price Index (IPCA), which is collected at the monthly frequency. IPCA inflation is a key variable for the *Brazilian Inflation-Targeting Regime*, since it is the official inflation-target variable.

If the number of survey respondents were very large, this could potentially serve to approximate a large N, T environment, since our data covers forecasts collected every working day from the period of January 2nd, 2006, to February 7th, 2014 ($2,020$ working days)[11]. However, since we must rely (with respect to nowcasts) on the active group of about 100 institutions to comprise the cross-section, the framework here is that of a large T with fixed N, or, at best, one in which we let first $T \to \infty$, and then let $N \to \infty$.

In every working day considered within our sample, market agents, $i = 1, ..., N$, may inform their expectations regarding inflation rates all the way up to the next 14 months. The raw data contains forecasts for "fixed-events" and varying forecast horizons; see Bakhshi et al. (2005). To fit the setup discussed in (18) and (19), the original forecasts are re-organized to form time-series of fixed horizons h and time-varying events ($t = 1, ..., T = 98$ months). As a result, the data base forms an unbalanced panel ($N \times T \times H$) containing an amount of $2,732,827$ observations. The final data base used in this paper contains $1,486,559$ observations.

Decomposing our total of $1,486,559$ observations into N, T, and H, gives the following breakdown: $t = 1, ..., T = 98$ months (or events), $h = 1, ..., H = 400$ days (or

[11] The *Focus Survey* database has been collected since 1999, the year in which the Brazilian inflation-targeting regime was created. However, in its inception, the survey had only a small coverage (small N), that being the reason why we start the forecasting exercise in 2006.

forecast horizons), and an average of $i = 1, ..., N = 37.9$ forecasters, when considering the full term-structure of 14 monthly forecasts. We must also stress that the great majority of survey respondents provided nowcasts and/or short-term forecasts, but only a smaller set of respondents inform their forecasts for the full term-structure of forecasts, up to the longest horizon.

Despite not fitting exactly the large N and T environment of *big data*, the forecasting gains we report below are non-trivial, and can serve as a benchmark to the lower bound of gains expected to be present on a big-data environment. To implement our forecasting exercise, we split our time-series sample into two consecutive sub-periods: the first ($t = 1, ..., T_1$) is labeled as "training sample", where realizations of y_t are usually confronted with forecasts provided by the survey, and potential bias-correction terms are estimated using either (18) or (19). We choose $T_1 = 60$ months. The second sub-period is where genuine out-of-sample forecasts are entertained. This period comprises the last P observations of our sample ($t = T_1+1, ..., T$) – where $P = T-T_1 = 38$ months. These P observations are thus used to compare different forecasts, computing forecast-accuracy measures.[12]

To evaluate a given forecast *method*, we 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 the h-step ahead forecast (of y_t formed at period $t - h$) of any given method. Here, we considered five types of forecast methods. The first is the one proposed in this paper – which we have labelled the *extended bias-corrected average forecast (EBCAF)*. The second is the simple cross-sectional average forecast – *average forecast (AF)* for short. It has a long tradition in econometrics, all the way from Granger (1969), to Stock and Watson (2006), and Capistrán and Timmermann (2009), among others. The third is the *bias-corrected average forecast* – *BCAF* for short. It was proposed by Issler and Lima (2009) and performs an intercept correction of the average forecast. The fourth is the *Principal Component Analysis* (PCA), which is a well-known data-reduction technique, applied here to summarize the information content of the cross-section of forecasts; see Forni et al. (2000) and Stock and Watson (2002b). The last method is the $AR(1)$ model. Within the $ARMA$ class of models, it is the best one-step ahead predictor of Brazilian inflation for our sample, according to several different measures of Information Criteria. It will be used here as our basic benchmark.

In constructing the *EBCAF*, we used a set of instruments containing lagged inflation π_t and interest i_t rates (IPCA and SELIC, respectively). Interest rates are

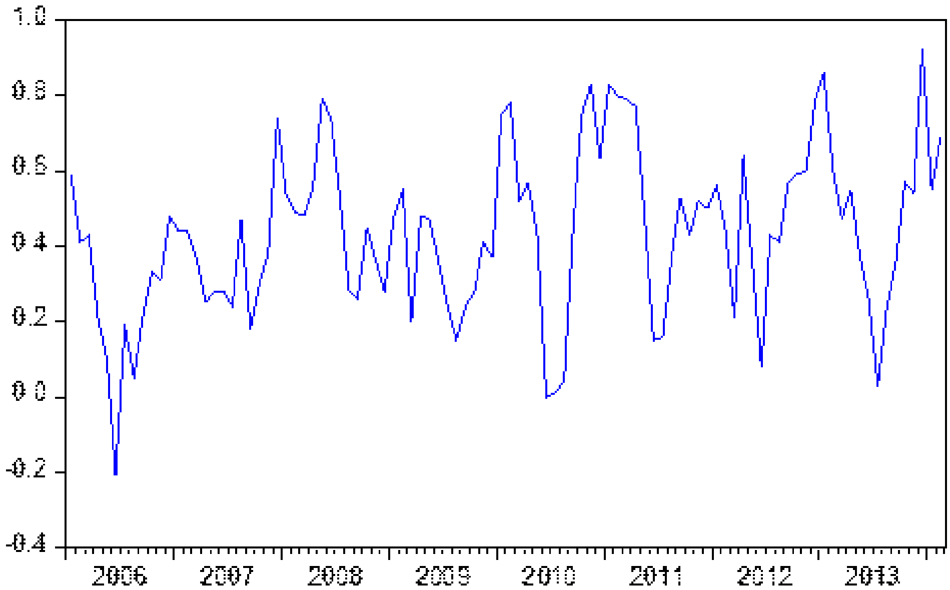
[12] We use a recursive estimation scheme (i.e., increasing training sample size). In this context, each model is initially estimated using the first $T_1 = 60$ observations (excepting the average forecast *method*) and the out-of-sample point forecasts of y_{T_1+h}, $h = 1, ..., H$, are generated. We, then, add an additional observation at the end of the training sample, re-estimate the models and generate again out-of-sample forecasts. This process is repeated along the remaining data.

transformed by first-differences of logs. The results are based on the following set of instruments: $z_{t-s} = [1; \pi_{t-s}; \pi_{t-s-2}; \pi_{t-s-5}; \Delta \ln i_{t-s-5}]'$ with $s = 14$ months for the longest horizon and $s = 1$ month for the shortest. In GMM estimation, the key averages of parameters in the extended BCAF were estimated by using (18) for each horizon (h), instead of using its stacked version – equation (19). In practice, estimation of the stacked version (19) was not feasible. The empirical exercise was conducted using the *R software* (version R i386 3.2.2), and the package "GMM" encoding the "two step" approach of Hansen (1982) was employed, although the "iterative" procedure of Hansen et al. (1996) yielded very similar results.

3.2 Empirical Results

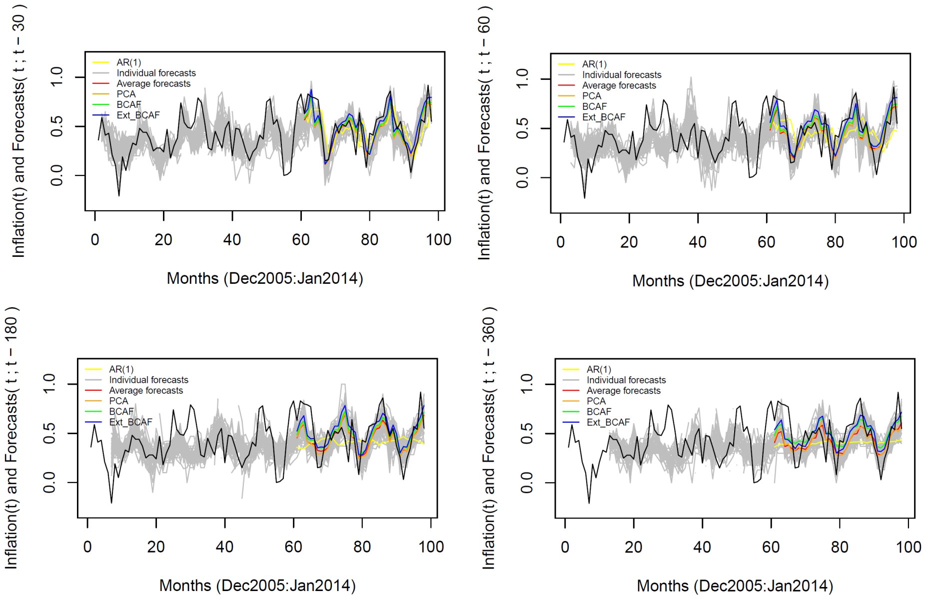
The results of our empirical exercise are next presented. Figure 1 shows the monthly CPI inflation (IPCA) in Brazil. Inflation seems to conform to a stationary-ergodic process. For example, all tests for unit roots reject $I(1)-ness$. As noted before, within the $ARMA$ class, inflation is best described by an $AR(1)$ model with an estimated $AR(1)$ coefficient of 0.621 – estimated standard error of 0.068. Figure 2 plots monthly inflation and the respective daily forecast for survey respondents at horizons $h = 30, 60, 180, 360$ days. It also presents the $AR(1)$ forecasts, the average forecasts, the PCA[13] forecasts, the BCAF, and the extended BCAF, at these same horizons.

Figure 1 - Inflation rate (y_t) in Brazil: % monthly



[13] First principal component of the panel of forecasts, computed recursively along the out-of-sample forecasting exercise. For $h = 1$, 6 and 12 months, it accounts, respectively, for 84%, 73% and 68% of total variation contained in the cross-section of forecasts.

Figure 2 - Inflation rate y_t and inflation forecasts for selected horizons ($h = 30$, 60, 180 and 360 days, respectively)



Notes: Yellow lines present the AR(1) forecasts. Gray lines show the forecasts $f_{i,t}^{h}$ of survey participant i for y_t made at period $t-h$. Red represent the average forecast, orange denotes the PCA forecast, and the green and blue lines show, respectively, the BCAF and the Extended BCAF forecasts. Black line is the inflation rate y_t.

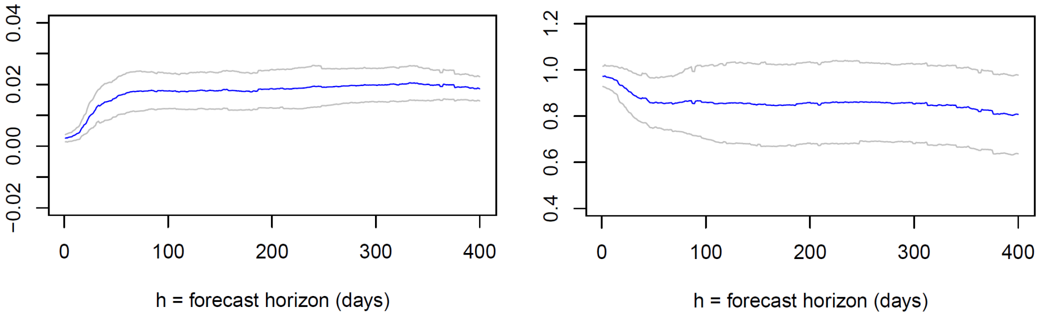
Table 1 reports GMM estimates for the extended BCAF. Although estimates of the average intercept are close to zero, and estimates of the average slope are close to unity, they are all highly significantly different to zero at any given horizon. A Wald-test for zero average intercept and unity average slope highly rejects the null at all horizons, showing the usefulness of the approach proposed in this paper. Figure 3 plots these two estimates across daily horizons with 95% confidence bands.

Table 1 - GMM estimation results

horizon h (days)	$\widehat{\overline{k^h}}$	$\widehat{\overline{\beta^h}}$	Wald test (p-value)	OIR test (p-value)
10	0.0036 (0.0009)	0.9638 (0.0276)	$1.1E-13$	0.7059
20	0.0071 (0.0016)	0.9312 (0.0399)	$3.0E-12$	0.7038
30	0.0118 (0.0026)	0.8929 (0.0481)	$7.5E-16$	0.6237
60	0.0173 (0.0033)	0.8541 (0.0583)	$9.6E-11$	0.6532
90	0.0180 (0.003)	0.8647 (0.0763)	$1.8E-12$	0.7095
180	0.0176 (0.003)	0.8486 (0.0884)	$2.8E-11$	0.6884
360	0.0199 (0.0025)	0.8275 (0.088)	$2.9E-15$	0.5893

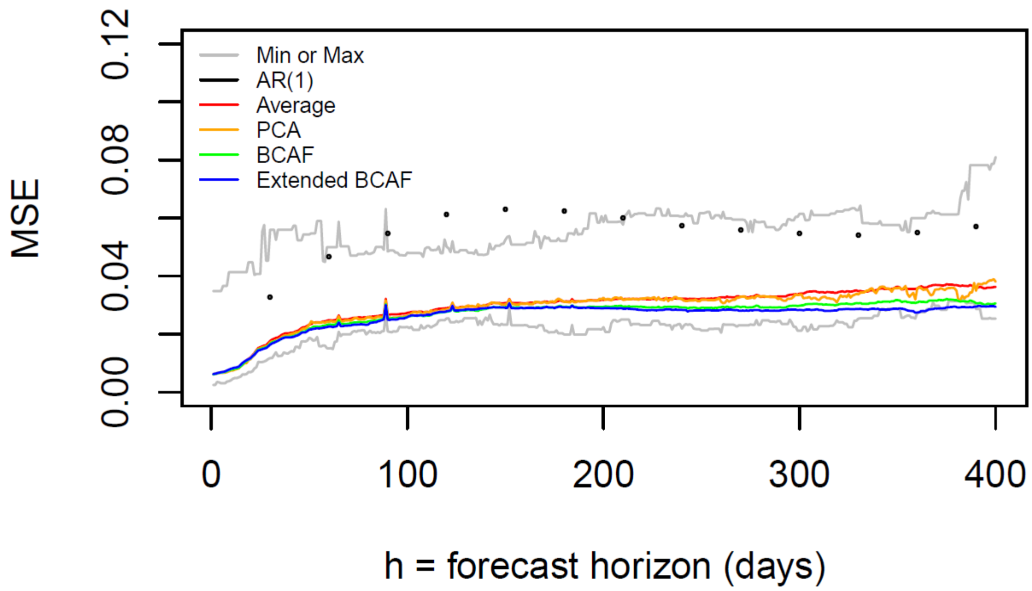
Notes: Robust standard errors in parentheses. Wald test refers to $Ho: [k^h; \beta^h] = [0; 1]$. Last column shows the p-value of the Over-Identifying Restriction (OIR) J-test due to Hansen (1982).

Figure 3 - Estimates $\widehat{k^h}$ (left panel) and $\widehat{\beta^h}$ (right panel) and 95% confidence intervals



We compare the out-of-sample MSE of each of the five forecast methods employed here: the AF, the PCA, the BCAF, the extended BCAF and the $AR(1)$ model as well as the minimum and maximum MSEs across all survey respondents. Figure 4 plots the daily results for these methods[14]. The first interesting feature is the comparison between the $AR(1)$ model and the forecast-combination techniques discussed here – the AF, the PCA, the BCAF, and the extended BCAF. Strikingly, notice that the $AR(1)$ model is usually close to upper bound of the MSE, while the other four methods are all close to the lower bound. This shows that forecast combination works in practice – in our case, the optimal combination of survey forecasts. These are exactly the results obtained by Ang et al. (2007) and Faust and Wright (2013).

Figure 4 - Mean Squared Error (MSE)



Note: Max (Min) denotes the maximum (minimum) MSE, for each horizon, across all forecasters.

Next, when we compare the MSE of the five methods discussed above there is a *pecking order*: the extended BCAF is superior to the BCAF, which in turn dominates the AF (and the PCA), which in turn dominates the $AR(1)$ model. This validates

[14] Recall that the AR(1) model is estimated at the monthly frequency.

the view that the bias corrections performed either by the BCAF, or by the extended BCAF, are a useful device for forecasting using surveys. As shown in Table 2 and Figure 5 (left panel), when we compare the extended BCAF, or the BCAF, with the AF or the PCA, forecasting MSE reductions can reach up to 25% at some horizons for the extended BCAF, and can reach up to 15% for the BCAF, although they are usually at the 10%-15% range. Reductions vis-à-vis the MSE associated with forecasts of the $AR(1)$ model can reach up to 50% for the extended BCAF.

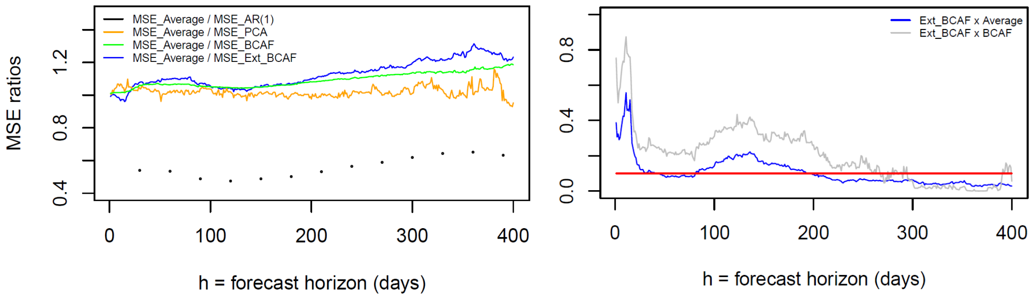
Table 2 - Mean Squared Error (MSE)

horizon h (days)	AR(1)	PCA	Average Forecast	BCAF	Extended BCAF
30	0.0328* [0.000]	0.0170 [0.400]	0.0177 (0.105)	0.0166 (0.247)	0.0164
60	0.0475* [0.001]	0.0242 [0.348]	0.0248* (0.081)	0.0234 (0.206)	0.0226
90	0.0523* [0.005]	0.0257 [0.414]	0.0267 (0.119)	0.0254 (0.261)	0.0249
180	0.0466* [0.049]	0.0311 [0.359]	0.0312 (0.124)	0.0294 (0.297)	0.0293
270	0.0415* [0.046]	0.0329 [0.108]	0.0328* (0.056)	0.0293* (0.099)	0.0281
360	0.0452* [0.019]	0.0356* [0.029]	0.0357* (0.028)	0.0309* (0.000)	0.0275

Notes: The second and third columns show [in brackets] the p-values of the test of Diebold-Mariano (1995) which compares the Extended BCAF with the forecast in each column. The other columns show (in parenthesis) the p-values of the test of Clark and West (2007), which compares the Extended BCAF and the forecast in each column. In all cases, * indicates a rejection of the null at a 10% level.

We test whether or not forecast errors differ in a statistical sense. Figure 5 (right panel) presents the equal-predictive-accuracy test of Clark and West (2007) for nested models at different horizons. Results suggest that, vis-à-vis the average forecast, both bias-correction devices (BCAF and the extended BCAF) can statistically reduce (at the 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). Comparisons of the extended BCAF with the $AR(1)$ model using a Diebold-Mariano (DM) test for equal variances shows that the former is statistically superior to the latter (at the 10% significance level) at all horizons. The DM test also indicates that the extended BCAF is statistically superior to the PCA-based forecast at the one-year horizon.

Figure 5 - MSE comparison (left panel) and
Clark and West (2007) test (p-values, right panel)



Note: On the right panel, the red line represents a p-value of 0.10. Ho: equal predictive accuracy.

Regarding heterogeneous beliefs, Proposition 5 allows the decomposition of the estimated EBCAF into public and private terms. If all the information contained in the variables of the public information set $\mathcal{F}_{t-h}$ could be summarized by a standard dynamic factor model, we can disentangle $\mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}\right)+\mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right)$ into two orthogonal components. To do so, we consider the factor model: $y_t = \Lambda f_t + \sum_{j=1}^{12} \delta_j d_{j,t}^{seas} + \epsilon_t$, where f_t are the unobserved factors, derived from a set of 95 macroeconomic and financial variables (see Table A.1 in the Appendix) using Principal Component Analysis[15], and $d_{j,t}^{seas}$ are seasonal dummies. Following the literature (e.g., Bánbura et al., 2013) we also specify the factors as following a vector autoregressive (VAR) process $f_t = \Phi(L) f_t + u_t$.

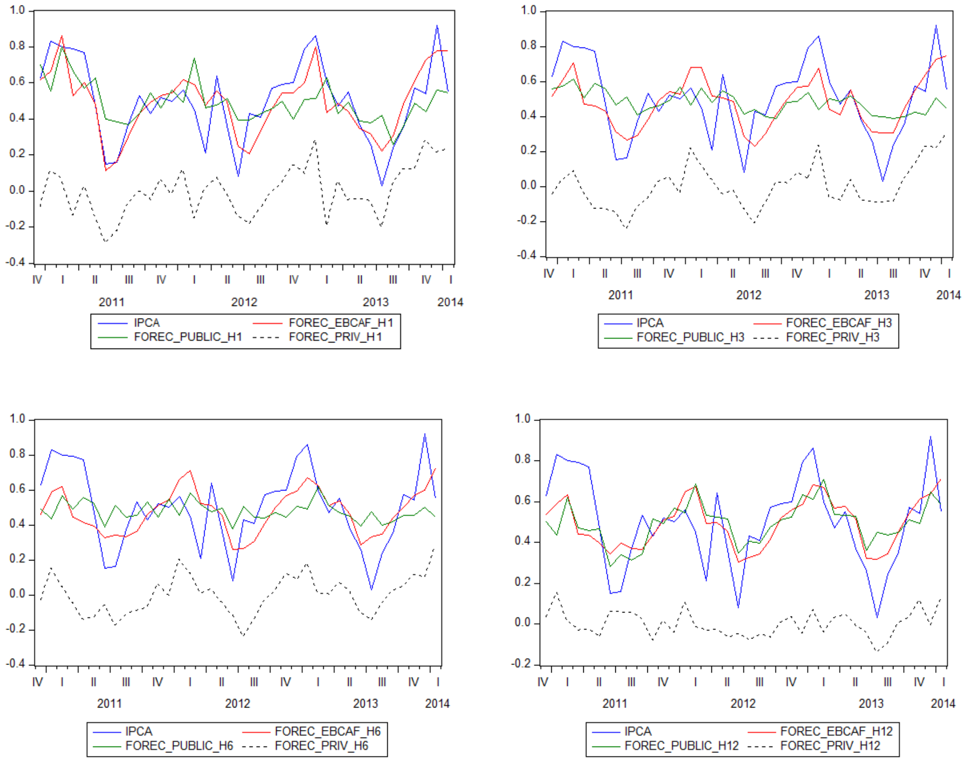
This way, $\widehat{\mathbb{E}}\left(y_t \mid \mathcal{F}_{t-h}\right)$ is the projection of $\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{k^h}}{\widehat{\beta^h}}$ onto $\widehat{\Lambda}\widehat{f}_{t|t-h} + \sum_{j=1}^{12} \widehat{\delta}_j d_{j,t}^{seas}$, where $\widehat{\Lambda}$ and $\widehat{\delta}_j$ are consistent estimates of Λ and δ_j, and $\widehat{f}_{t|t-h}$ is the updated factor consistent estimate.[16] $\widehat{\mathbb{E}}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right)$ is the corresponding residual of that projection. Figure 6 presents the results for selected horizons h. Note that the public information series is highly correlated with the estimated EBCAF (e.g., 0.63 for $h = 1$ month, and 0.84 for $h = 12$ months).

In turn, the private information series (for $h = 1, 3, 6$ and 12 months) does not Granger cause, at 5% level of significance, the interest rate (SELIC), the country risk-premium (Embi+Br) or the stock market index (Ibovespa) monthly variation. This makes sense, since the residual should represent the mean across agents of all the idiosyncratic expectation components. While we should expect individual idiosyncratic expectation components to be important for some agents and for some periods, the importance should vanish when taking into account the average across the market.

[15] From the set of 95 variables, we extract the first 12 principal components (which jointly accounts for 70% of the total variance in the data) to form a monthly VAR with 1 lag (following the Hannan-Quinn information criterion).

[16] It is the $h-$step ahead out-of-sample forecast from the VAR (estimated in a recursive scheme).

Figure 6 - Forecast decomposition (public *versus* private) for selected horizons ($h = 1, 3, 6$ and 12 months)



4 Conclusion

This paper proposes a *financial approach* to economic forecasting which can be applied to data bases of surveys of forecasts. From a forecasting perspective, the focus on surveys is important, since Faust and Wright (2013) have shown forcefully that subjective forecasts of inflation seem to outperform model-based forecasts. We model the forecasting decision of an individual from *first principles* (i.e., *microfounded*) and show that surveys of forecasts obey an affine factor structure with a single factor which is the common component of the conditional expectation of the target variable. This holds in a context where individuals only have access to public information or where they also have access to private information with common and idiosyncratic components.

Our approach involves two layers of decision making. In the fist, individuals choose the best forecast to be posted in the database containing surveys of forecasts. In the second, an *econometrician* uses survey results to forecast the target variable optimally. From the point of view of the econometrician, it is important to estimate the common factor of the survey, which is the optimal forecast under public information and an *MSE risk function*. We show how this can be performed using GMM-based estimates. In this context, the feasible optimal forecast is a function of the consensus forecast of the survey (a cross-sectional average of survey forecasts) after appropriately filtering out two bias terms. This links the *financial approach* of economic forecasting to the forecast-combination literature, where idiosyncratic risk of individual forecasts can be diversified out.

Our results are applicable to two types of surveys with a large enough number of time observations ($T \to \infty$): one of *current surveys of forecasts*, which possess a limited number of respondents ($N < \infty$), and one that we have labelled as the *Surveys of the Future*, where the number of respondents is also large ($N \to \infty$). The latter connects this paper with the literature on *big data*. In standard GMM moment estimation, we circumvent the *curse of dimensionality* that arises from the factor structure (large N) by employing cross-sectional averages. In a big-data context, this allows the use of all the information contained in the survey, while estimating a parsimonious factor model, with only two biases – intercept and slope bias.

We apply the techniques advanced in this paper to forecast Brazilian inflation using the *Focus Survey*, organized by the Central Bank of Brazil. It is a unique panel database, which collects daily information from 254 registered *professional institutions*, which are followed throughout time with a reasonable turnover. There is a smaller *active* group of about 100 institutions that frequently update their forecasts. Forecasts are supplied at the daily frequency for a large array of macroeconomic time-series at different forecast horizons. Our sample covers forecasts for inflation rates from December 2005 to January 2014 ($t = 1, ..., T = 98$ months), and the forecast horizons range from $h = 1, ..., H = 400$ days. The final data base used in this paper contains $1,486,559$ observations, forming an unbalanced panel of agents.

We compare the truly out-of-sample mean-squared error (MSE) of five different forecast methods – four of them widely used in the forecasting literature: the method proposed here – the *extended* bias-corrected average forecast (extended BCAF), the average forecast (AF, the cross sectional average of forecasts), the principal-component analysis (PCA), the bias-corrected average forecast (BCAF), and the forecast of the $AR(1)$ model for inflation, which is the best $ARMA$ model in sample using the Schwarz Criterion. There is *pecking order*: the extended BCAF is superior to the BCAF, which in turn dominates the AF (and the PCA), which in turn dominates the $AR(1)$ model. When we compare the extended BCAF with the AF, MSE reductions can reach up to 25% at some horizons. Reductions vis-à-vis the MSE associated with forecasts of the $AR(1)$ model can reach up to 50%.

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] Apostol, T.M., 1967. *Calculus, Vol. 1: One-Variable Calculus with an Introduction to Linear Algebra* (2nd Ed.), New York: John Wiley & Sons.

[3] Athanasopoulos, G., Guillén, O.T.C., Issler, J.V., Vahid, F., 2011. Model Selection, Estimation and Forecasting in VAR Models with Short-run and Long-run Restrictions. *Journal of Econometrics* 164 (1), 116-129.

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

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

[6] Bańbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow. Working Paper Series n.1564, European Central Bank.

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

[8] Capistrán, C., Timmermann, A., 2009. Forecast Combination with Entry and Exit of Experts. *Journal of Business and Economic Statistics* 27, 428-440.

[9] Christoffersen, P.F., Diebold, F.X., 1997. Optimal prediction under asymmetric loss. *Econometric Theory* 13, 808-817.

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

[11] 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.

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

[13] Diebold, F.X., 2012. A Personal Perspective on the Origin(s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline. Manuscript, University of Pennsylvania, http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

[14] Diebold, F.X., Mariano, R.S., 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13, 253-263.

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

[16] Einav, L., Levin, J.D., 2014. The Data Revolution and Economic Analysis, in, *Innovation Policy and the Economy*, Josh Lerner and Scott Stern, editors. NBER Book: University of Chicago Press.

[17] 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.

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

[19] Elliott, G., Timmermann, A., 2005. Optimal forecast combination weights under regime switching. *International Economic Review* 46 (4), 1081-1102.

[20] Engle, R.F., Kozicki, S., 1993. Testing for Common Features (with comments), *Journal of Business and Economic Statistics* 11, 369-395.

[21] Faust, J., Wright, J.H., 2013. Forecasting Inflation. Handbook of Economic Forecasting, Volume 2A, Chapter 1, p.3-56. Ed. Elsevier B.V.

[22] Forni, M., Hallim, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics* 82, 540 554.

[23] Forni, M., Hallim, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* 100, 830 840.

[24] Gaglianone, W.P., Issler, J.V., Matos, S.M., 2017. Applying a microfounded-forecasting approach to predict Brazilian inflation. *Empirical Economics* 53 (1), 137-163.

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

[26] Gaglianone, W.P., Lima, L.R., 2014. Constructing Optimal Density Forecasts From Point Forecast Combinations. *Journal of Applied Econometrics* 29 (5), 736-757.

[27] Gaglianone, W.P., Lima, L.R., Linton, O., Smith, D.R., 2011. Evaluating Value-at-Risk Models via Quantile Regression. *Journal of Business and Economic Statistics* 29 (1), 150-160.

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

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

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

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

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

[33] 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.

[34] Issler, J.V., Vahid. F., 2001. Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates. *Journal of Monetary Economics* 47 (3), 449-475.

[35] Issler, J.V., Vahid, F., 2006. The missing link: Using the NBER recession indicator to construct coincident and leading indices of economic activity. *Journal of Econometrics* 132, 281-303.

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

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

[38] Lahiri, K., Peng, H., Sheng, X., 2015. Measuring Uncertainty of a Combined Forecast and a New Test for Forecaster Homogeneity. Mimeo: University at Albany, SUNY.

[39] Marques A.B.C., 2013, Central Bank of Brazil's market expectations system: a tool for monetary policy. *Bank for International Settlements*, Volume 36, 304-324.

[40] Mincer, J.A., Zarnowitz, V., 1969. The Evaluation of Economic Forecasts. In: Jacob Mincer, Ed., Economic Forecasts and Expectations. New York: National Bureau of Economic Research.

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

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

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

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

[45] Stock, J., Watson, M., 2002a. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics* 20, 147-162.

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

[47] Stock, J., Watson, M., 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.

[48] 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.

[49] Vahid, F., Engle, R.F., 1993. Common Trends and Common Cycles. *Journal of Applied Econometrics* 8, 341-360.

[50] Vahid, F., Engle, R.F., 1997. Codependent cycles. *Journal of Econometrics* 80, 199-221.

[51] Vahid, F., Issler, J.V., 2002. The importance of common cyclical features in VAR analysis: A Monte Carlo study. *Journal of Econometrics* 109, 341–363.

[52] Varian, H.R., 2014. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives* 28 (2), 3-28.

Appendix: Proofs of Propositions

Proof of Proposition 1. (i) Parts of this proof follows 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 given by A5, the optimal forecast can be represented in the following way:

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

$$\begin{aligned}
&= \underset{\widetilde{y}}{\arg\min} \int L^i\left(\frac{1}{(\gamma_0+\gamma_1 x_{t,t-h})}(y-\widetilde{y})\right)dF_{t,t-h}(y) \\
&= \underset{\widetilde{y}}{\arg\min} \int L^i\left(\frac{1}{(\gamma_0+\gamma_1 x_{t,t-h})}(\delta_0+\delta_1 x_{t,t-h}+\gamma_0\eta_t+\gamma_1 x_{t,t-h}\eta_t-\widetilde{y})\right)dF_{\eta,h}(\eta).
\end{aligned}$$

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:

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

in which we have used the fact that $F_{\eta,h}(\eta)$ is time-invariant by definition, and $\gamma^i_h \equiv \underset{\widetilde{\gamma}}{\arg\min} \int L^i(\eta_t-\widetilde{\gamma})\,dF_{\eta,h}$ 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 quantile $\tau_i \in [0,1]$, so that $\widetilde{f}^h_{i,t} = \alpha_0(\tau_i)+\alpha_1(\tau_i)x_{t,t-h}$. On the other hand, 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; see Koenker (2005, p.302). 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 $F_{\eta,h}(0,1)$ and on the loss function L^i.

(ii) 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 2. From Proposition 1, 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 $\{y_t; x_{t,t-h}\}^T_{t=1}$. Now, in order to prove consistency of $[\widehat{k^h_i}; \widehat{\beta^h_i}]$, define $\widehat{\alpha}(\tau) \equiv [\widehat{\alpha}_0(\tau); \widehat{\alpha}_1(\tau)]$. Koenker (2005, Theorem 4.1, p.120) shows that the estimator $\widehat{\alpha}_n(\tau) = \underset{\alpha\in\mathbb{R}^2}{\arg\min} \sum^n_{t=1} \rho_\tau(y_t - \alpha_0 - \alpha_1 x_{t,t-h})$, 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 quantiles as long as $n \to \infty$ (under conditions A1-A2 of Koenker (2005, p.120)). This way, it follows that $\widehat{\overline{\alpha}} = \sum^K_{k=1} \widehat{\alpha}(\tau_k)\Delta\tau_k \xrightarrow{p} \sum^K_{k=1} \alpha(\tau_k)\Delta\tau_k$, provided that $\widehat{\alpha}(\tau) \xrightarrow{p} \alpha(\tau)$ for all $\tau \in [0,1]$. Now, applying the second fundamental theorem of calculus (or the Newton–Leibniz axiom) on the later sum, it follows that $\lim_{K\to\infty} \left(\sum^K_{k=1} \alpha(\tau_k)\Delta\tau_k\right) = \int_0^1 \alpha(\tau)d\tau = \overline{\alpha}$, where a Riemann integral is obtained in the limit $K \to \infty$ (see Apostol, 1967) and the partitions $\Delta\tau_k = \frac{1}{K+1}$ get finer (i.e., $\Delta\tau_k \to 0$ as long as $K \to \infty$). Therefore, it follows that $\widehat{k^h_i} \equiv \left(\widehat{\alpha}_0(\tau_i) - \frac{\widehat{\overline{\alpha_0}}}{\widehat{\overline{\alpha_1}}}\widehat{\alpha}_1(\tau_i)\right) \xrightarrow{p}$ $k^h_i = \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}}\alpha_1(\tau_i)\right)$ and $\widehat{\beta^h_i} \equiv \frac{\widehat{\alpha}_1(\tau_i)}{\widehat{\overline{\alpha_1}}} \xrightarrow{p} \beta^h_i = \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}$. ■

Proof of Proposition 3. Assumptions A1-A6 and Proposition 2 are necessary to justify an optimal feasible forecast of y_t conditioned on $\mathcal{F}_{t-h}$ of the form $f^h_{i,t} = k^h_i + \beta^h_i\mathbb{E}_{t-h}(y_t) + \varepsilon^h_{i,t}$. Then, assumptions A7-A9 are used to achieve the set of orthogonality conditions (18) or (19), which are used to estimate the set of parameters $\theta^h = [k^h; \beta^h]'$ within a T-consistent GMM context. To show that, first consider the affine function $f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h = k_i^h + \beta_i^h y_t + \beta_i^h \eta_t^h + \varepsilon_{i,t}^h$. Taking the cross-sectional average and letting $N \to \infty$, it follows (from A8) that $f_{\cdot,t}^h = k^h + \beta^h y_t + \beta^h \eta_t^h + \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t}^h$. Now, evaluate the previous equation at respective GMM estimates $\widehat{\theta^h} = [\widehat{k^h}; \widehat{\beta^h}]'$, such that: $f_{\cdot,t}^h = \widehat{k^h} + \widehat{\beta^h} y_t + \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h}$, where $\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h} = \widehat{\beta^h}\widehat{\eta_t^h} + \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \widehat{\varepsilon_{i,t}^h}$. Notice that also letting $T \to \infty$ in the previous equation, it follows that $\operatorname*{plim}_{(N,T\to\infty)_{\text{seq}}} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} v_{i,t}^h = \beta^h \eta_t^h + \operatorname*{plim}_{(N,T\to\infty)_{\text{seq}}} \frac{1}{N} \sum_{i=1}^{N} \widehat{\varepsilon_{i,t}^h} = \beta^h \eta_t^h + \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t}^h = \beta^h \eta_t^h$, since from A7 and Lemma 1 of Issler and Lima (2009), it follows that $\operatorname*{plim}_{N\to\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t}^h \right) = 0$. Thus, $\operatorname*{plim}_{(N,T\to\infty)_{\text{seq}}} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h} = \beta^h \eta_t^h$. On the other hand, recall that $\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h} = f_{\cdot,t}^h - \widehat{k^h} - \widehat{\beta^h} y_t$ and, thus, $\left(\frac{f_{\cdot,t}^h - \widehat{k^h}}{\widehat{\beta^h}} \right) = y_t + \frac{\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h}}{\widehat{\beta^h}}$.

Now, letting $T \to \infty$, it follows that $\operatorname*{plim}_{(N,T\to\infty)_{\text{seq}}} \left(\frac{f_{\cdot,t}^h - \widehat{k^h}}{\widehat{\beta^h}} \right) = y_t + \operatorname*{plim}_{T\to\infty} \left(\frac{\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h}}{\widehat{\beta^h}} \right)$

$= y_t + \frac{\operatorname*{plim}_{(N,T\to\infty)_{\text{seq}}} \frac{1}{N} \sum_{i=1}^{N} \widehat{v_{i,t}^h}}{\operatorname*{plim}_{T\to\infty} \left(\widehat{\beta^h} \right)} = y_t + \frac{\beta^h \eta_t^h}{\beta^h} = \mathbb{E}_{t-h}(y_t)$, since (from A8-A9) it follows that $\operatorname*{plim}_{T\to\infty} \left(\widehat{\beta^h} \right) = \beta^h = \beta_0^h \neq 0$, where (for notation purposes) we used $\theta^h = [k^h; \beta^h]'$ to denote the populational parameters (of the microfounded model for the optimal forecast) and $\theta_0^h = [k_0^h; \beta_0^h]'$ as the unique set of values that solves (within a GMM setup) either (18) or (19) for each h separately. ■

Proof of Proposition 4. From A10, and based on the definition of η_t^h, it follows that $\mathbb{E}_{t-h}\left(\varepsilon_{i,t}^h\right) = \mathbb{E}_{t-h}\left(\eta_t^h\right) = 0$. Thus, for a given survey respondent i and respective set of parameters $\theta_i^h = [k_i^h; \beta_i^h]'$ it follows that $\mathbb{E}_{t-h}\left(f_{i,t}^h - k_i^h - \beta_i^h y_t\right) = \mathbb{E}_{t-h}\left(\beta_i^h \eta_t^h + \varepsilon_{i,t}^h\right) = \beta_i^h \mathbb{E}_{t-h}\left(\eta_t^h\right) + \mathbb{E}_{t-h}\left(\varepsilon_{i,t}^h\right) = 0$ and, thus, $\mathbb{E}_{t-h}\left(y_t\right) = \left(\frac{f_{i,t}^h - k_i^h}{\beta_i^h} \right)$, for all i, t and given h. It also follows that $\mathbb{E}_{t-h}\left(y_t\right) = \frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - k_i^h}{\beta_i^h} = \frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}}$, where the first equality is simply the cross-section average applied to the equation $\mathbb{E}_{t-h}\left(y_t\right) = \left(\frac{f_{i,t}^h - k_i^h}{\beta_i^h} \right)$, and the second equality comes from A11 (definition of average parameters $\overline{\theta^h} = [\overline{k^h}; \overline{\beta^h}]'$). Now, evaluate the previous equation at respective GMM estimates $\widehat{\overline{\theta^h}} = \left[\widehat{\overline{k^h}}; \widehat{\overline{\beta^h}} \right]'$, such that: $\widehat{\mathbb{E}}_{t-h}(y_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{\widehat{k^h}}}{\widehat{\widehat{\beta^h}}}$. Then, let $T \to \infty$ to obtain $\underset{T\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{\widehat{k^h}}}{\widehat{\widehat{\beta^h}}}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}} = \mathbb{E}_{t-h}(y_t)$, due to A11 (consistency of $\widehat{\widehat{\theta^h}} = \left[\widehat{\widehat{k^h}}; \widehat{\widehat{\beta^h}}\right]'$ estimates for all N). In addition, letting (sequentially) $N \to \infty$ does not change the convergence to the conditional mean, that is, $\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{\widehat{k^h}}}{\widehat{\widehat{\beta^h}}}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}} = \mathbb{E}_{t-h}(y_t)$, provided that (also from A11), for all N, there is a unique set of values $\overline{\theta_0^h} = [\overline{k_0^h}; \overline{\beta_0^h}]'$ that solves either (18) or (19) for each h separately. ■

Proof of Proposition 5. Following the proof of Proposition 2, and using $\mathbb{E}_{i,t-h}(y_t)$ for a given individual i instead of $\mathbb{E}_{t-h}(y_t)$, it follows that $f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{i,t-h}(y_t) + \varepsilon_{i,t}^h$. Now consider the orthogonal decomposition $y_t = \mathbb{E}_{i,t-h}(y_t) - \eta_{i,t}^h$, where $\mathbb{E}_{i,t-h}(\eta_{i,t}^h) = 0$. Thus, $f_{i,t}^h = k_i^h + \beta_i^h y_t + \beta_i^h \eta_{i,t}^h + \varepsilon_{i,t}^h$. Taking the cross-sectional average and letting $N \to \infty$, following the proof of Proposition 3, it follows (from A8 and A12) that: $f_{\cdot,t}^h = k^h + \beta^h y_t + \beta^h \eta_t^h + \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h\right)$. Now, evaluate the previous equation at GMM estimates $[\widehat{k^h}; \widehat{\beta^h}]'$, such that: $f_{\cdot,t}^h = \widehat{k^h} + \widehat{\beta^h} y_t + \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h}$, where $\underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h} = \widehat{\beta^h}\widehat{\eta_t^h} + \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{\varepsilon_{i,t}^h}$. Also letting $T \to \infty$ in the previous equation, it follows that: $\underset{(N,T\to\infty)_{\text{seq}}}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h} = \beta^h \eta_t^h + \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h = \beta^h \eta_t^h$, since from A7 and Lemma 1 of Issler and Lima (2009): $\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h\right) = 0$. Thus, $\underset{(N,T\to\infty)_{\text{seq}}}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h} = \beta^h \eta_t^h$. On the other hand, recall that: $\underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h} = f_{\cdot,t}^h - \widehat{k^h} - \widehat{\beta^h} y_t$ and, thus, $\left(\frac{f_{\cdot,t}^h - \widehat{k^h}}{\widehat{\beta^h}}\right) = y_t + \frac{\underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h}}{\widehat{\beta^h}}$. Now, letting $T \to \infty$, it follows that $\underset{(N,T\to\infty)_{\text{seq}}}{\text{plim}}\left(\frac{f_{\cdot,t}^h - \widehat{k^h}}{\widehat{\beta^h}}\right) = y_t + \underset{T\to\infty}{\text{plim}}\left(\frac{\underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h}}{\widehat{\beta^h}}\right) = y_t + \frac{\underset{(N,T\to\infty)_{\text{seq}}}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h}}{\underset{T\to\infty}{\text{plim}}\left(\widehat{\beta^h}\right)} = y_t + \frac{\beta^h \eta_t^h}{\beta^h} = \widetilde{\mathbb{E}}_{t-h}(y_t)$. The last equality comes from A12, and applying the cross-sectional average to the orthogonal decomposition: $y_t = \mathbb{E}_{i,t-h}(y_t) - \eta_{i,t}^h \therefore y_t = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{i,t-h}(y_t) - \frac{1}{N}\sum_{i=1}^{N}\eta_{i,t}^h$, also letting $N \to \infty$ as follows: $y_t = \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{i,t-h}(y_t)\right) - \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_{i,t}^h\right)$. Since $\mathcal{F}_{t-h}$, $\mathcal{F}_{t-h}^{PC}$ and $\mathcal{F}_{i,t-h}^{PI}$ form a *partition* of $\mathcal{F}_{i,t-h}$, from the properties of conditional expectation as an orthogonal projection operator, it follows that $\mathbb{E}_{i,t-h}(y_t) = \mathbb{E}[y_t | \mathcal{F}_{t-h}] + \mathbb{E}\left[y_t | \mathcal{F}_{t-h}^{PC}\right] + \mathbb{E}\left[y_t | \mathcal{F}_{i,t-h}^{PI}\right]$. By taking the cross-sectional average on both sides of the previous expression, and letting $N \to \infty$, it follows that: $\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}\right)\right)$ $+\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(y_t \mid \mathcal{F}_{t-h}^{PC}\right)\right) = \mathbb{E}\left[y_t \middle| \mathcal{F}_{t-h}\right] + \mathbb{E}\left[y_t \middle| \mathcal{F}_{t-h}^{PC}\right]$, since from A12: $\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{PI}\right)\right) = 0.$ ■

Table A.1

Category	Name	Category	Name
Prices	IPCA (consumer price index)	Exterior	Exports (FOB, total)
Prices	IPCA (consumer price index, market prices)	Exterior	Exports (FOB, primary goods)
Prices	IPCA (consumer price index, regulated and monitored prices)	Exterior	International reserves (total)
Prices	IPC-Fipe (consumer price index)	Exterior	Current account (monthly, net)
Prices	IPC-Br (consumer price index)	Exterior	Current account (accumulated in 12 months, in relation to GDP)
Prices	IPA-DI (wholesale price index)	Exterior	FDI (Foreign Direct Investment, accumulated in 12 months)
Prices	IGP-DI (general price index)	Exterior	FPI (Foreign Portfolio Investment, accumulated in 12 months)
Prices	IGP-M (general price index)	Econ. activity	IBC-BR (central bank economic activity index)
Money	Monetary base	Econ. activity	GDP (accumulated in the last 12 months, current prices)
Money	Money supply (currency outside banks)	Labor	Unemployment rate (open)
Money	Demand deposits	Labor	Registered employees index (wholesale and retail trade)
Money	Savings deposits	Labor	Registered employees index (construction sector)
Money	M1	Labor	Hours worked in production (São Paulo)
Money	M2	Labor	Real overall wages (industry, São Paulo)
Money	M3	Industry	Industrial production (total)
Money	M4	Industry	Industrial production (mineral extraction)
Interest rates	Nominal policy interest rate (Selic)	Industry	Industrial production (manufacturing industry)
Interest rates	Nominal policy interest rate (long-term interest rate, TJLP)	Industry	Industrial production (capital goods)
Interest rates	Nominal market interest rate (Swap Pré-DI, 1 year)	Industry	Industrial production (intermediate goods)
Interest rates	Real market interest rate (Swap Pré-DI, 1 year, Focus 12m inflation expect.)	Industry	Industrial production (consumer goods)
Banking sector	Credit spread (nonearmarked credit rate - Selic rate)	Industry	Industrial production (durable goods)
Banking sector	Non-Performing Loans (NPL) of total credit	Industry	Industrial production (semidurable and nondurable goods)
Banking sector	Loan-to-Deposit ratio (LTD)	Industry	Installed capacity utilization (São Paulo)
Banking sector	Reserve requirements ratio (financial inst. reserve requir./ total deposits)	Industry	Capacity utilization (manufacturing industry, FGV)
Banking sector	Real growth of nonearmarked credit operations outstanding	Industry	Vehicles production (total)
Capital markets	Ibovespa (Brazil)	Industry	Passenger cars and light commercial vehicles production
Capital markets	Initial Public Offers (IPOs) accumulated in 12 months (Brazil)	Industry	Truck production
Capital markets	Net equity of stock funds (Brazil)	Industry	Bus production
Capital markets	Net equity of financial investment funds (Brazil)	Industry	Production of agricultural machinery (total)
Capital markets	MSCI emerging countries (EM, US$)	Sales	Sales volume index in the retail sector (total)
Capital markets	MSCI developed countries (World, US$)	Sales	Sales volume index in the retail sector (fuel and lubricants)
Risk premium	Embi+Br (Emerging Markets Bond Index Plus Brazil, spread)	Sales	Sales volume index in the retail sector (superm., food, bever. and tobacco)
Risk premium	Embi+composite (average spread of 16 emerging countries)	Sales	Sales volume index in the retail sector (textiles, clothing and footwear)
Risk premium	CDS (Credit Default Swap, Brazil 5 years)	Sales	Sales volume index in the retail sector (furniture and white goods)
Exchange rates	FX-rate (nominal exchange rate, R$/US$)	Sales	Sales volume index in the retail sector (vehicles and motorcycles, spare parts)
Exchange rates	REER (Real effective exchange rate, IPA-13 currencies)	Sales	Sales volume index in the retail sector (hypermarkets and supermarkets)
Global Economy	U.S. dollar index (DXY, geometric average of 6 currencies in respect to US$)	Sales	Vehicle sales (total)
Global Economy	U.S. Treasury 2 years (Treasury nominal interest rates)	Sales	Domestic vehicle sales
Global Economy	U.S. Treasury 10 years (Treasury nominal interest rates)	Energy	Electric energy consumption (commercial)
Global Economy	U.S. Treasury 5 years TIPS (Treasury Inflation-Protected Securities)	Energy	Electric energy consumption (residential)
Global Economy	CRB all commodities index	Energy	Electric energy consumption (industrial)
Global Economy	Oil price (WTI, Oklahoma-USA)	Energy	Electric energy consumption (other)
Global Economy	VIX CBOE volatility index (30-day expected volatility of the S&P500)	Energy	Electric energy consumption (total)
Exterior	Import price index	Public sector	Net public debt (% GDP, total, federal government and central bank)
Exterior	Import quantum index	Public sector	Net public debt (% GDP, internal, federal government and central bank)
Exterior	Export price index	Public sector	Net public debt (% GDP, external, federal government and central bank)
Exterior	Export quantum index	Public sector	Primary result of consolidated public sector (flow accum. in 12 months, % GDP)
Exterior	Imports (FOB, total)		