Microfounded Forecasting*

Wagner Piazza Gaglianone†
João Victor Issler‡

June 18, 2018

Abstract

Our focus is on information in expectation surveys that can now be built on thousands (or millions) of respondents on an almost continuous-time basis (big data) and in continuous macroeconomic surveys with a limited number of respondents. We show that, under standard microeconomic and econometric techniques, survey forecasts are an affine function of the conditional expectation of the target variable. This is true whether or not the survey respondent knows the data-generating process (DGP) of the target variable or the econometrician knows the respondent's individual loss function. If the econometrician has a mean-squared-error risk function, we show that asymptotically efficient forecasts of the target variable can be built using Hansen's (*Econometrica*, 1982) generalized method of moments in a panel-data context, when N and T diverge or when T diverges with N fixed. Sequential asymptotic results are obtained using Phillips and Moon's (*Econometrica*, 1999) framework. Possible extensions are also discussed.

Keywords: Forecast Combination, Big Data, Common Features, Panel Data.
JEL Classification: C14, C33, E37.

*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. Also, all data manipulations were executed by employees of the Banco Central do Brasil. We are especially grateful for the comments and suggestions given by Scott Atkinson, Robert Engle, Antonio Galvão, Raffaella Giacomini, Alain Hecq, Bo Honoré, Fabian Krüger, Marcelo Moreira, Whitney Newey, Andrew Patton, Rafael Santos and Mark Watson. 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); and The Conference of the International Association for Applied Econometrics (London). Any remaining errors are ours. 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

1 Introduction

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. Along the same lines, 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.

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. Of course, there is already a large literature on *big data* in statistics. It is certainly true that econometrics is currently benefiting from it. However, what we see as lacking is interpretation – economic interpretation. One step in that direction is to employ the tools of economics to be able to extend our current knowledge in an emerging field. This is one of the objectives of this paper.

In order to accomplish that, our focus is on the abundant information contained in expectation surveys that can now be built on thousands (or millions) of respondents on an almost continuous-time basis. 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 users of that information, it is important to us to consider an environment with three different setups: (i) when the survey respondent knows the data-generating process (DGP) of the response random variable, where the DGP may take an analytic parametric form, and users know what is the loss function being used by the respondent; (ii) when the respondent knows the DGP, where the DGP is driven by a location-scale model with known parameters, and users know what is the loss function being used; and (iii) when the respondent does not know the parameters of the DGP, the latter being approximated using a location-scale model and quantile regression techniques, and the loss function being used by respondent is unknown to the users. Under these conditions, we show that the optimal forecast of a given survey respondent is an affine function of the conditional expectation of the target variable of the survey. Once we characterize this first layer of decision making, we can then ask how to use efficiently the informational content present in expectation surveys.

Our answer uses structural microfounded models, whereas the previous literature had focused on a reduced-form approach.

If the final user of that information (the second layer of decision making) 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., then, it is natural to extract from the survey the conditional expectation of the target variable[1] – the optimal forecast from the point-of-view of the final user, e.g., an econometrician. Because we had an affine model for the optimal respondent's decision, the second-layer problem reduces to a signal-extraction problem, where potentially the number of respondents and of time-series observations is large, i.e., N and T grow without bound. In this context, we show how to solve this problem using Hansen's (1982) generalized method of moments (GMM) in the sequential asymptotic framework of Phillips and Moon (1999), i.e., our method identifies and proposes consistent and efficient estimates of the optimal forecast of the target variable under the MSE risk function. Moreover, we also propose a solution for the case in which T diverges but N is fixed, which applies to some long standing surveys with a fixed number of respondents throughout time.[2]

Although the affine model of the first-layer decision making is a function of a latent variable (the conditional expectation), we are able to substitute it for the target variable itself minus a martingale-difference sequence (MDS) error term, thus expressing the cross-sectional moment condition in terms of this MDS, where its unforecastable properties can then be exploited using GMM. So, the moment condition can ultimately be cast only in terms of observables and parameters. In this setting, we confront forecasts with realizations to be able to identify and estimate consistently functions of the parameters of the affine model when the number of time-series observations gets large – large T. Our results hold in two different contexts regarding the number of survey responses: when they are large ($N \to \infty$) and when N is fixed.

Obviously, our primary focus is on forecasting. As Einav and Levin (2013) 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. While business analytics is a big deal and surely has improved the efficiency of many organizations, predictive modeling lies behind many striking information products and services introduced in recent years." However, since our techniques allow the identification and estimation of the conditional expectation of the target variable, it immediately allows for the identification of economic shocks to these target variables, which are key variables especially in macroeconomics and finance; see the discussion in Diebold (2012).

[1] Of course, different risk functions in the second layer would imply extracting a different target. This is not a problem here, given that we have properly characterized the first-layer decisions.

[2] For example, the *Livingston* Survey of the Philadelphia FED, the *Wall Street Journal* Survey of Forecasters, and the *Survey of Professional Forecasters*.

We should stress that we avoid dealing with sparse-data issues in *big data*, e.g., Belloni and Chernozhukov (2011, 2013) and Belloni, Chernozhukov, and Wang (2011), because our focus narrows down to the relationship between optimal forecasts and the conditional expectation of the target variable in the survey. Implicitly, each survey respondent could use several different covariates to compute their optimal forecast. Indeed, all of these may be unknown to the user in the second layer of decision making (the *econometrician*). Because the number of respondents diverge, our techniques benefits from the information implicit in all of these forecasts, and could ultimately be interpreted as an optimal forecast-combination method. Thus, our approach to *big data* is not related to data reduction techniques – where the key issue is the availability of too many regressors (Varian, 2014) – but on how to optimally combine information.

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), 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 (2013). From the theoretical literature on the incentives leading to optimal forecasts, our paper is related to the work of Laster et al. (1999), Ottaviani and Sorensen (2006), and Batchelor (2007). 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), and to the novel literature on informational rigidities, e.g., Mankiw and Reis (2010), Coibion and Gorodnichenko (2012), and Andrade and Le Bihan (2013). From *big data* econometrics it is related to the work of Diebold (2012), Einav and Levin (2013), Varian (2014), Warin and Sanger (2014), and Joseph et al. (2014), although we could not find papers dealing directly on how to optimally combine forecasts in the context of big data. Indeed, most of this literature is interested in dimension-reduction techniques for potential regressors, such as classification and regression trees (CART), random forests, penalized-regression techniques such as the least absolute shrinkage and selection operator (LASSO) and least-angle regression (LARS), and elastic nets.

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 a survey of Brazilian inflation expectations using the methods proposed here, comparing the out-of-sample performance of different bias-correction approaches. Section 4 concludes.

2 Econometric Setup

2.1 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 variable in question. 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 $f^h_{i,t}$, $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. Therefore, $f^h_{i,t}$ are h-step-ahead forecasts of y_t, formed at period $t - h$, and N is the number of respondents of this opinion poll 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]:

$$f^h_{i,t} = \beta^h_i \mathbb{E}_{t-h}(y_t) + k^h_i + \varepsilon^h_{i,t}. \tag{1}$$

As is well known, Granger (1969) was a pioneer in this 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 term as in $\mathbb{E}_{t-h}(y_t) + k^h_i$.

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

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

of a random variable y_t by using a specific *loss function* under different assumptions about knowledge of the DGP of y_t. These optimal forecasts $f_{i,t}^h$ will be available as survey results, where the number of respondents is potentially large, i.e., $N \to \infty$, and these surveys can be periodically taken on a large number of different occasions, i.e., $T \to \infty$. This setup describes reasonably well the current abundance of knowledge in 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 cases of potential asymmetries or unknown loss functions previously studied are taken into account. We investigate the relationship between the optimal forecast and the conditional expectation, based on the following cases: (1) Loss function known to the econometrician and DGP known by individual i, where the DGP may take an analytic parametric form; (2) Loss function known to the econometrician and DGP known by individual i, where the DGP is approximated by a location-scale model with known parameters; (3) Loss unknown to the econometrician and parameters of DGP unknown by individual i, the latter being estimated by the individual using standard quantile-regression techniques and a location-scale model.

Loss and DGP known – (Cases (1) and (2))

Assume that there is an amount of individuals forecasting y_t conditional on $\mathcal{F}_{t-h}$, where $\mathcal{F}_{t-h}$ is the sigma algebra of all the variables contained in the conditioning set used by all the agents. Each individual chooses an optimal forecast $\widetilde{f}_{i,t}^h$ by minimizing their respective expected loss function L^i, e.g., Granger and Newbold (1986). In this section, we assume that the agent in the second layer of decision making – the *econometrician* – has full knowledge of the Loss function used by individual agents. So, there is full disclosure not only of survey-forecast results but also of the Loss function from the part of survey respondents. This last assumption will be relaxed in the next section.

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

[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.

Therefore, the optimal (point) individual forecasts of y_t are obtained as follows:

$$\widetilde{f}_{i,t}^{h} \equiv \arg\min_{f} \ \mathbb{E}\left(L^i(y_t; f) \mid \mathcal{F}_{t-h}\right) \tag{2}$$

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 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 (2) has no constraints regarding its cost. These features will have bearing when we discuss our theoretical results vis-à-vis those of the macroeconomics literature on forecasting.

A natural assumption about the shape of the agent's loss function 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 class of loss functions. A simple way to consider an asymmetric function, and account for some "degree of asymmetry", is given by the following assumption:

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 $F^{-1}_{t,t-h}(\tau)$, $\forall \tau \in (0;1)$, 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.

Proposition 1 *(Asymmetric Loss and Density) Assume A1-A4. (i) If* $\beta_1 > \beta_2$ *and* $\mathbb{E}\left(\left|g^{i\prime}(e^i_{t,t-h})\right|\right) \leq K$, *where* $K \equiv \int_{-\infty}^{\widetilde{f}^h_{i,t}} g^{i\prime}(e_i) f_{t,t-h}(y)dy \Big/ \int_{-\infty}^{\widetilde{f}^h_{i,t}} f_{t,t-h}(y)dy$, *and* $f_{t,t-h}$ *is a unimodal and asymmetric conditional density of* y_t, *such that* $\mathbb{E}_{t-h}(y_t) > Med_{t-h}(y_t)$, *where* $Med_{t-h}(y_t)$ *is the conditional median of* y_t, *then* $\widetilde{f}^h_{i,t} < \mathbb{E}_{t-h}(y_t)$. *Also, for two forecasters* i *and* j *such that* $\beta^i_1/\beta^i_2 > \beta^j_1/\beta^j_2 > 1$ *it follows that* $\widetilde{f}^h_{i,t} < \widetilde{f}^h_{j,t} < \mathbb{E}_{t-h}(y_t)$. *(ii) If* $\beta_1 < \beta_2$, $\mathbb{E}\left(\left|g^{i\prime}(e^i_{t,t-h})\right|\right) \geq K$ *and* $f_{t,t-h}$ *is unimodal and asymmetric, such that* $\mathbb{E}_{t-h}(y_t) < Med_{t-h}(y_t)$, *then* $\widetilde{f}^h_{i,t} > \mathbb{E}_{t-h}(y_t)$. *Also, for two forecasters* i *and* j *such that* $\beta^i_1/\beta^i_2 < \beta^j_1/\beta^j_2 < 1$ *it follows that* $\widetilde{f}^h_{i,t} > \widetilde{f}^h_{j,t} > \mathbb{E}_{t-h}(y_t)$.

Proposition 2 *(Asymmetric Loss and Symmetric Density) Assume A1-A4 and that* $f_{t,t-h}$ *is unimodal and symmetric. (i) If* $\beta_1 > \beta_2$, *then* $\widetilde{f}^h_{i,t} < \mathbb{E}_{t-h}(y_t)$. *(ii) If* $\beta_1 < \beta_2$, *then* $\widetilde{f}^h_{i,t} > \mathbb{E}_{t-h}(y_t)$.

Proposition 3 *(MSE Loss, Granger, 1969) Assume the Mean Squared Error loss function. If* $|\mathbb{E}_{t-h}(y_t)| < \infty$ *for all* $t = 1, ..., T$ *and* $h = 1, ..., H$, *then* $\widetilde{f}^h_{i,t} = \mathbb{E}_{t-h}(y_t)$.

Proposition 1 states two important results: First, asymmetries in both the loss and the DGP generate departures of the optimal forecast from the central tendency (e.g., mean) of the conditional distribution of y_t. Second, for the same DGP, the higher the degree of asymmetry in the loss function the greater will be the distance between the optimal forecast and the conditional mean. From Proposition 2, note that the symmetry of the conditional density of y_t is not a sufficient condition to ensure the conditional mean to be the optimal forecast. According to Proposition 3, under the MSE loss, and assuming the existence of the conditional expectation of y_t for all t and h, the optimal forecast is the conditional mean of y_t, which is the well-known result of the conditional mean optimality under MSE risk function, due to Granger (1969).

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

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

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

Granger (1969) argues that symmetry of both the loss function and the conditional density of y_t is not a sufficient condition for the optimum predictor to be equal to the conditional mean. In fact, he provides a counter-example in which the conditional mean would be sub-optimal under symmetric functions (both loss and the PDF). In order to better understand the relationship between the optimal forecast and the conditional mean, we next assume specific parametric versions for the DGP, as follows:

Proposition 5 *(DGP - parametric PDFs) If A1-A4 hold and the conditional PDF of* y_t *is: (i) Gaussian, Two-piece Normal, or Logistic, then,* $\widetilde{f}^h_{i,t} = k^h_i + \mathbb{E}_{t-h}(y_t)$*; (ii) Log-normal or Weibull, then,* $\widetilde{f}^h_{i,t} = \beta^h_i \mathbb{E}_{t-h}(y_t)$*; (iii) Beta*$(a = 1, b > 0)$*, then,* $\widetilde{f}^h_{i,t} = k^h_i + \beta^h \mathbb{E}_{t-h}(y_t)$*; (iv) Beta*$(a > 0, b = 1)$*, then,* $\widetilde{f}^h_{i,t} = \beta^h_i \varphi(\mathbb{E}_{t-h}(y_t); \tau_i)$*, where* $\varphi(\mathbb{E}_{t-h}(y_t); \tau_i) = \exp(\frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)})$ *and* $\tau_i \equiv F_{t,t-h}(\widetilde{f}^h_{i,t})$*.*

With the exception of case (iv) above, it is instructive to note that Proposition 5 delivers an encompassing result that:

$$\widetilde{f}^h_{i,t} = \beta^h_i \mathbb{E}_{t-h}(y_t) + k^h_i, \tag{3}$$

where different parametric CDFs deliver different restrictions on either intercepts and/or slopes of the affine function (3). Different intercepts and slopes are usually a result of the interaction between the DGP and the Loss function, in some cases being linked to asymmetry of the Loss function. If both are asymmetric, as in (iv) above, we do not obtain an affine function, but a nonlinear factor model on $\mathbb{E}_{t-h}(y_t)$.

Knowledge of the Loss function (by the econometrician) and the DGP (by the agent) is too stringent an assumption, making it of little practical interest. Thus, we

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

now turn into a more realistic case where we assume a location-scale model for the DGP of y_t. This broad class of DGPs was used by Patton and Timmermann (2007), among others. Several parametric DGPs belong to it, such as the Normal distribution, Elliptical, Cauchy, Uniform, Logistic, Student's t, and Generalized extreme value. The class also includes most common volatility processes, such as ARCH and stochastic volatility. It is not only broad but also very convenient, because it can naturally be investigated through the lens of the quantile-regression framework with linear conditional quantiles.[6]

Assumption A5 (DGP - location-scale) The DGP of y_t follows a location-scale model, with conditional mean and variance dynamics defined as $y_t = X'_{t,t-h}\delta + \left(X'_{t,t-h}\gamma\right)\eta_t$, in which $(\eta_t|\mathcal{F}_{t-h}) \sim i.i.d.\ F_{\eta,h}(0,1)$, where $F_{\eta,h}(0,1)$ is some distribution with zero mean and unit variance, which depends on h but does not depend on $\mathcal{F}_{t-h}$; $X_{t,t-h} \in \mathcal{F}_{t-h}$ is a $m \times 1$ vector of covariates (which includes the intercept, and that can be predicted using information available at time $t-h$) and $\delta = [\delta_0; \delta_1; ...; \delta_{m-1}]$ and $\gamma = [\gamma_0; \gamma_1; ...; \gamma_{m-1}]$ are $m \times 1$ vectors of parameters.[7]

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

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

Here, scale effects generate heterogeneity on the slopes of the affine structure. On the other hand, the intercept is a function of both scale and location parameters. We now turn into the case where the Loss function is unknown by the econometrician. Moreover, the DGP is also unknown by the agent and the econometrician. However, it can be approximated using the location-scale model by the agent in forming her/his

[6] The linear quantile regression setup could be further extended to consider models with dependency and mixing conditions (e.g. Cai and Xiao, 2012; Galvao and Wang, 2015).

[7] 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 Appendix A.3 for some possible ways to consider the multicovariates case.

optimal forecast. In other words, the parameters δ and γ, associated with the location-scale model, can be consistently estimated as we shall see next.

Loss unknown and parameters of DGP unknown (Case (3))

So far, we have assumed that the Loss function is known by the econometrician and that the DGP of y_t is known to the survey respondent who is forecasting 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. In this context, we consider the case where the econometrician only observes a survey of individual forecasts and the target variable y_t, but has no information at all about the individual loss functions previously used in the first layer of decision making. Moreover, in the first layer, the survey respondent does not know the DGP and has to estimate the parameters of the location-scale approximation.

Now, an optimal forecast of the form:

$$\widetilde{f}_{i,t}^{h} = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t), \tag{4}$$

obtained in the previous section, is not feasible. Despite that, because the class of location-scale models is relatively broad, we assume that the unknown DGP falls within this class (A5) and that survey respondents employ the location-scale model to approximate the unknown DGP, which parameters must then be estimated.

A natural way for survey respondents to estimate (4) 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 estimators of k_i^h, β_i^h, and $\mathbb{E}_{t-h}(y_t)$, the optimal (feasible) survey responses are:

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

given the assumption that survey respondents use the location-scale model as an approximation.

Notice that we have made the distinction between the optimal forecast $\widetilde{f}_{i,t}^{h}$ and its 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 (4) versus (5). We now use a technical assumption that guarantees the existence of quantile-regression consistent estimates.

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 7 *If A1-A6 hold, then, the optimal (feasible) forecast of y_t conditioned on $\mathcal{F}_{t-h}$ is of the form: $f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h$, where $\varepsilon_{i,t}^h$ accounts for finite sample parameter uncertainty, and $[\widehat{k_i^h};\widehat{\beta_i^h}]$ are consistent estimates of $[k_i^h;\beta_i^h]$.*

Notice that the case of unknown Loss function and DGP is similar to the two previous cases. Adding and subtracting $k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t)$ to (5), leads to:

$$f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h, \tag{6}$$

where it becomes clear that $\varepsilon_{i,t}^h = \left[\widehat{k_i^h} - k_i^h\right] + \left[\widehat{\beta_i^h}\widehat{\mathbb{E}}_{t-h}(y_t) - \beta_i^h \mathbb{E}_{t-h}(y_t)\right]$ reflects the small-sample error in approximating the unknown DGP[8]. Indeed, with one exception, (6) encompasses all cases covered above, which are special cases reflecting specific parameter restrictions. Next, we will employ the encompassing model (6) as a starting point to propose a way to identify and estimate consistently (and efficiently in a GMM context) the random variable $\mathbb{E}_{t-h}(y_t)$.

As is clear from above, quantile-regression techniques allowed the estimation of $\widehat{\mathbb{E}}_{t-h}(y_t)$, $\widehat{k_i^h}$ and $\widehat{\beta_i^h}$, 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. The 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 (6). 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 forecasts are functions of $\mathcal{F}_{t-h}$, 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.

[8] As it is clear from the text, an optimal forecast "approximation error" $\varepsilon_{i,t}^h \equiv f_{i,t}^h - \widetilde{f}_{i,t}^h$ arises in (6). In the Appendix A.2, we provide a full account of these error terms using as examples a few parametric DGPs.

Identification and GMM estimation of $\mathbb{E}_{t-h}(y_t)$

Consider an econometrician who only observes a survey of individual forecasts $f_{i,t}^h$ (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[9]. From our previous results, it is natural for the econometrician to use:

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

As argued above, this encompasses the previous cases of Loss and DGP known, and of Loss and DGP unknown[10]. Equation (7) 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 rapidly converge to their unconditional counterparts, making the case for large H unappealing. So, our framework entails two interesting cases. The first is $T \to \infty$, $N \to \infty$, and small H, which applies to large surveys taken across time[11]. The second 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.

Even if H is small, the system of equations (7) has too many parameters when N is large, which poses a problem for identification of $\mathbb{E}_{t-h}(y_t)$. However, 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 (7) across i, 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 as function of three means alone:

$$\mathbb{E}_{t-h}(y_t) = \operatorname*{plim}_{N\to\infty} \left(\frac{\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i^h}{\frac{1}{N}\sum_{i=1}^{N} \beta_i^h} \right), \tag{8}$$

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

[10] There is one exception noted above.

[11] Examples include large cross-sectional repeated surveys, where individuals are re-sampled with a small probability, allowing the construction of pseudo-panels (or synthetic panels) of individuals.

as long as all the terms in parenthesis in (8) 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 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 $\mathbb{E}_{t-h}(\eta_t^h) = 0$, so that:

$$y_t = \mathbb{E}_{t-h}(y_t) - \eta_t^h. \tag{9}$$

Thus, combining (7) and (9) leads to:

$$\begin{aligned} f_{i,t}^h &= k_i^h + \beta_i^h(y_t + \eta_t^h) + \varepsilon_{i,t}^h & (10)\\ &= k_i^h + \beta_i^h y_t + v_{i,t}^h, & (11) \end{aligned}$$

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 reasonable to expect that the approximation error $\varepsilon_{i,t}^h$ vanishes.

Starting with $\mathbb{E}\left(v_{i,t}^h \mid \mathcal{F}_{t-h}\right) = 0$, 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:

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

which is valid for all $i = 1, ..., N$, $t = 1, ..., T$, and $h = 1, ..., H$. The system (12) 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 (12) 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 parameter dimensionality[12], leading to:

$$\mathbb{E}\left[\left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \otimes z_{t-s}\right] = 0, \tag{13}$$

$t = 1, ..., T$, 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 (13) 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. 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 (13) is to stack all the restrictions across h (finite) as:

$$\mathbb{E}\left[\begin{pmatrix} \overline{f_{\cdot,t}^1} - \overline{k^1} - \overline{\beta^1} y_t \\ \overline{f_{\cdot,t}^2} - \overline{k^2} - \overline{\beta^2} y_t \\ \vdots \\ \overline{f_{\cdot,t}^H} - \overline{k^H} - \overline{\beta^H} y_t \end{pmatrix} \otimes z_{t-s}\right] = 0, \tag{14}$$

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 $dim(z_{t-s}) > 2$. Given a choice of H, GMM estimation of (14) is efficient. A less efficient alternative to estimate the whole stacked system (14) is to estimate separately (13) 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 less stringent 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 (13) and (14) would converge in probability to a unique limit as $N \to \infty$, i.e., $\operatorname{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \beta_i^h = \beta^h$, $\beta^h \neq 0$, $\left|\beta^h\right| < \infty$, $\operatorname{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} k_i^h = k^h$, $\left|k^h\right| < \infty$, and $\operatorname{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h = f_{\cdot,t}^h$, $\left|f_{\cdot,t}^h\right| < \infty$, for all $t = 1, 2, \cdots, T$. Then, a key condition for

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

T-consistent estimation is that (13) and (14), evaluated at these limits are a unique solution for each of them respectively. Take now moments (13), considering that $N \to \infty$:

$$\mathbb{E}\left[\left(f_{.,t}^{h} - k^{h} - \beta^{h} y_{t}\right) \otimes z_{t-s}\right] = 0, \text{ or,} \tag{15}$$

$$\mathbb{E}\left[\left(\beta^{h}\eta_{t}^{h} + \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^{h}\right) \otimes z_{t-s}\right] = 0. \tag{16}$$

If, in addition, $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^{h} \xrightarrow{p} 0$, as $N \to \infty$, (16) collapses to:

$$\mathbb{E}\left[\beta^{h}\eta_{t}^{h} \otimes z_{t-s}\right] = 0, \tag{17}$$

$t = 1, ..., T$, and $h = 1, ..., H$, where (17) 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 $\boldsymbol{\varepsilon}_{t}^{h} = \left(\varepsilon_{1,t}^{h}, \varepsilon_{2,t}^{h}, ... \, \varepsilon_{N,t}^{h}\right)'$ 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

$$\lim_{N\to\infty}\frac{1}{N^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\mathbb{E}\left(\varepsilon_{i,t}^{h}\varepsilon_{j,s}^{h}\right)\right| = 0, \tag{18}$$

for all t and s, given h.

Assumption A8 We assume that $\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\beta_{i}^{h} = \beta^{h} \neq 0$, $\left|\beta^{h}\right| < \infty$, $\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}k_{i}^{h} = k^{h}$, $\left|k^{h}\right| < \infty$, and $\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h} = f_{.,t}^{h}$, $\left|f_{.,t}^{h}\right| < \infty$, for all $t = 1, 2, \cdots, T$.

Assumption A9 We assume that the identification conditions for GMM estimation are met and that there is a unique set of values $\theta_{0}^{h} = [k_{0}^{h}; \beta_{0}^{h}]'$, $h = 1, 2, \cdots, H$, that solve either (14) or (13) for each h separately[13]. 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.

[13] Since the restrictions are linear, the θ_{0}^{h}, $h = 1, 2, \cdots, H$, are identified up to nonzero scalar multiplication.

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 (13) or (14), 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 8 *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 more 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 (13) or (14), but N is not large, we have to guarantee that all possible ensemble averages are 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 (13) and (14) 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 (13) and (14). Thus, we employ the following assumptions outlined next:

Assumption A10 Let $\boldsymbol{\varepsilon}_t^h=\left(\varepsilon_{1,t}^h,\varepsilon_{2,t}^h,...\,\varepsilon_{N,t}^h\right)'$ 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 $\mathbb{E}(\boldsymbol{\varepsilon}_t^h\mid\mathcal{F}_{t-h})=0$ 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 $\overline{\theta_0^h} = [\overline{k_0^h}; \overline{\beta_0^h}]'$, $h = 1, 2, \cdots, H$, that solves either (14) or (13) 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 $T \to \infty$ and then $N \to \infty$, or, that we let $T \to \infty$ with N fixed.

Proposition 9 *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.

So far, under double asymptotics, we discussed the sequential convergence approach, either $(N, T \to \infty)_{\text{seq}}$ or $(T, N \to \infty)_{\text{seq}}$, for the Extended BCAF estimator. Next, we ask under which conditions we can establish a link between sequential convergence and joint convergence. Following Phillips and Moon (1999), we state the formal conditions for the sequential convergence to imply joint convergence in our setup.

Assumption A12 Let first hold N, T, and h fixed, and define $Y_1 = \left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right)$. Letting now $T \to \infty$, this defines $\underset{T\to\infty}{\text{plim}} Y_1 \equiv Y_2 = \left(\frac{\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} \overline{k_i^h}}{\frac{1}{N}\sum_{i=1}^{N} \overline{\beta_i^h}}\right)$. If both $N, T \to \infty$, this defines $\underset{N,T\to\infty}{\text{plim}}\, Y_1 \equiv Y_3 = y_t + \eta_t^h = \mathbb{E}_{t-h}(y_t)$. We assume that Y_1, Y_2, and Y_3 are random variables on the same probability space $(\Omega, \mathcal{F}, P)$.

Assumption A13 $\underset{N,T}{\limsup} P\{\|Y_1 - Y_2\| > \varepsilon\} = 0, \forall \varepsilon > 0$, where $\|A\|$ is the Euclidean norm $(tr(A'A))^{1/2}$.

Proposition 10 *If A1-A13 hold, then, both feasible extended BCAFs* $\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^{h}-\widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right)$ *and* $\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^{h}-\widehat{k^h}}{\widehat{\beta^h}}\right)$*, based respectively on* T*-consistent GMM estimates* $\widehat{\overline{\theta^h}}=\left[\widehat{\overline{k^h}};\widehat{\overline{\beta^h}}\right]'$ *and* $\widehat{\theta}^h=\left[\widehat{k^h};\widehat{\beta^h}\right]$*, obey* $\underset{(T,N\to\infty)}{plim}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^{h}-\widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right)=\underset{(N,T\to\infty)}{plim}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^{h}-\widehat{k^h}}{\widehat{\beta^h}}\right)=\mathbb{E}_{t-h}(y_t)$*, regardless of the order in which* N *and* T *diverge.*

Under different assumptions, our results above imply that we can estimate consistently $\mathbb{E}_{t-h}(y_t)$, respectively, as follows:

$$\widehat{\mathbb{E}}_{t-h}(y_t)=\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^{h}-\widehat{k^h}}{\widehat{\beta^h}}\quad\text{or}\quad\widehat{\mathbb{E}}_{t-h}(y_t)=\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^{h}-\widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}},$$

depending on whether we let first $N\to\infty$, and then let $T\to\infty$, or, we either let first $T\to\infty$, and then let $N\to\infty$, or hold N fixed after $T\to\infty$. In any case, estimation of $\left[k^h,\beta^h\right]$ or of $\left[\overline{k^h},\overline{\beta^h}\right]$ is performed by GMM under T-asymptotics. These estimates of $\mathbb{E}_{t-h}(y_t)$ can be viewed as bias-corrected versions of survey forecasts. If the mean k^h $\left(\text{or }\overline{k^h}\right)$ is zero and the mean β^h $\left(\text{or }\overline{\beta^h}\right)$ is unity, these estimates converge to the same probability limit of the average (consensus) forecast $\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h}$, which has a long history in the forecasting literature.

Heterogeneous Beliefs

According to equation (3), each forecast is an affine function of the conditional expectation, and the informational content of each forecast over time is completely 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 several 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 in our setup, by allowing individuals to have different information sets $\mathcal{F}_{i,t-h}$ when forming their optimal (feasible) forecasts, that is:

$$f_{i,t}^{h}=k_i^h+\beta_i^h\mathbb{E}_{i,t-h}(y_t)+\varepsilon_{i,t}^{h},\tag{19}$$

where $\mathbb{E}_{i,t-h}(y_t)\equiv\mathbb{E}(y_t\mid\mathcal{F}_{i,t-h})$ and $\mathcal{F}_{i,t-h}$ is partitioned into three mutually exclusive (and non-empty) information sets, as follows: $\mathcal{F}_{i,t-h}=\mathcal{F}_{t-h}\cup\mathcal{F}_{i,t-h}^{PC}\cup\mathcal{F}_{i,t-h}^{PI}$ for

all $i = 1, ..., N$. The public information set $\mathcal{F}_{t-h}$ (i.e., the information available to all forecasters) is the subset of $\mathcal{F}_{i,t-h}$ whose elements are orthogonal to all the elements of the private information set $\left(\mathcal{F}^{PC}_{i,t-h} \cup \mathcal{F}^{PI}_{i,t-h}\right)$ which, in turn, is decomposed into the private-common information set $\mathcal{F}^{PC}_{i,t-h}$ (e.g., paid-up news/macro reports) and the private-idiosyncratic information set $\mathcal{F}^{PI}_{i,t-h}$ (e.g., private information, expert judgement). It is worth noting that the growth of N now implies a truly growth of information in the *big data*−sense. Thus, consider the following additional assumptions to be used in our result for heterogeneous beliefs, as summarized by Proposition 11.[14]

Assumption A14 Define $\widetilde{\mathbb{E}}_{t-h}(y_t) = \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{i,t-h}(y_t)\right)$, where $\left|\widetilde{\mathbb{E}}_{t-h}(y_t)\right| < \infty$ for all $t = 1, \cdots, T$ and $h = 1, ..., H$.

Assumption A15 Define $\eta^h_t = \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\eta^h_{i,t}\right)$ and $\beta^h\eta^h_t = \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\beta^h_i\eta^h_{i,t}\right)$, where $\left|\eta^h_t\right| < \infty$ and $\left|\beta^h\eta^h_t\right| < \infty$, for all $t = 1, \cdots, T$ and $h = 1, ..., H$.

Assumption A16 Let $\mathcal{F}_{t-h}$, $\mathcal{F}^{PC}_{i,t-h}$ and $\mathcal{F}^{PI}_{i,t-h}$ be mutually exclusive and non-empty information sets that form a partition of $\mathcal{F}_{i,t-h}$ for all $i = 1, ..., N$.

Assumption A17 Let $\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(y_t \mid \mathcal{F}^{PI}_{i,t-h}\right)\right) = 0$, for all $i = 1, ..., N$.

Assumption A18 Define $\mathbb{E}^{priv.com.}_{t-h}(y_t) = \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(y_t \mid \mathcal{F}^{PC}_{i,t-h}\right)\right)$ for all $i = 1, ..., N$ and $\mathbb{E}^{public}_{t-h}(y_t) = \mathbb{E}(y_t \mid \mathcal{F}_{t-h})$.

Proposition 11 *(Heterogeneous Beliefs) Under assumptions A1-A9 and A14-A18, the optimal (feasible) forecast of y_t conditioned on $\mathcal{F}_{i,t-h}$ is of the form: $f^h_{i,t} = k^h_i + \beta^h_i\mathbb{E}_{i,t-h}(y_t) + \varepsilon^h_{i,t}$, and the feasible Extended BCAF $\frac{1}{N}\sum_{i=1}^{N}\frac{f^h_{i,t}-\widehat{k^h}}{\widehat{\beta^h}}$, based on T-consistent GMM estimates $\widehat{\theta^h} = \left[\widehat{k^h};\widehat{\beta^h}\right]'$, obeys $\underset{(N,T\to\infty)_{\text{seq}}}{plim}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f^h_{i,t}-\widehat{k^h}}{\widehat{\beta^h}}\right) = \widetilde{\mathbb{E}}_{t-h}(y_t) = \mathbb{E}^{public}_{t-h}(y_t) + \mathbb{E}^{priv.com.}_{t-h}(y_t)$, 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 11 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, as follows: $\widetilde{\mathbb{E}}_{t-h}(y_t) = \mathbb{E}^{public}_{t-h}(y_t) + \mathbb{E}^{priv.com.}_{t-h}(y_t)$, where

[14] Proposition 11 provides an asymptotic result based on the sequential limits $(N, T\to\infty)_{\text{seq}}$. In order to save space, we do not provide the respective results for the cases of $(T, N\to\infty)_{\text{seq}}$ and $(N, T\to\infty)$, as covered in Propositions 9-10 in the homogeneous beliefs case.

$\mathbb{E}_{t-h}^{public}(y_t)$ is orthogonal to $\mathbb{E}_{t-h}^{priv.com.}(y_t)$ since $\mathcal{F}_{t-h} \cap \mathcal{F}_{i,t-h}^{PC} = \emptyset$. Moreover, if all the variables in the public information set $\mathcal{F}_{t-h}$ can be summarized by a factor model of the form: $y_t = c + \Lambda f_t + \varepsilon_t$, then one can recover the public information term $\widehat{\mathbb{E}}_{t-h}^{public}(y_t) = \widehat{c} + \widehat{\Lambda} f_{t,t-h}$, where $f_{t,t-h} = \mathbb{E}(f_t \mid \mathcal{F}_{t-h})$, as well as the private information term $\widehat{\mathbb{E}}_{t-h}^{priv.com.}(y_t) = \widehat{\widehat{\mathbb{E}}}_{t-h}(y_t) - \widehat{\mathbb{E}}_{t-h}^{public}(y_t)$. In our empirical exercise, we illustrate such possibility by deriving the factors f_t from a set of macroeconomic variables, using the PCA (Principal Component Analysis) data-reduction technique.

2.2 Discussion

When considering a survey of forecasts where the number of respondents and of time observations is large – *big data*, and also the case where the number of time observations is large but the number of respondents is fixed – standard continuous macroeconomic surveys, we apply the tools of the literature on optimal forecasts in the time-series dimension and model the forecasting decision of an individual from *first principles*. This constitutes the first layer of decision making. In our context, heterogeneity is a consequence of agents using different Loss functions and different information sets, which generate cross-sectionally diverse optimal forecasts. Despite that, each forecast is related to the conditional expectation of the target variable in the survey by an *affine* structure, where there is an intercept and slope bias due to the interaction between an asymmetric Loss and knowledge of the DGP. This affine factor setup can be exploited as a structural model for the conditional expectation.

Focusing on the conditional expectation defines the object of the second layer of decision making, when the user of the final information in the survey has an MSE risk function, e.g., central banks, the government, large firms, etc. The way we identify the conditional expectation can be viewed as a combination of cross-sectional averages with standard GMM moment conditions along the time dimension, where the affine structure offers natural orthogonality restrictions allowing the estimation of bias-correction terms. We circumvent the *curse of dimensionality* that arises from the factor structure (large N) by employing these cross-sectional averages. In a big-data context, this allows the use of all the information contained in the survey, while entailing the estimation of a parsimonious factor model.

Our setup is perfectly consistent with previous findings in the literature. For instance, our Proposition 6 can be directly linked to Proposition 3 of Patton and Timmermann (2007), which shows the optimal forecast as a conditional quantile of y_t at quantile level $\tau = q_h^*$. Since we assumed in A5 a location-scale model based on a vector of covariates $X_{t,t-h}$, affecting both the location and the scale simultaneously, it implies that the conditional quantiles of y_t are of the linear form: $\alpha_0(\tau) + \alpha_1(\tau) x_{t,t-h}$,

considering here[15] the simplest case $X_{t,t-h} = [1; x_{t,t-h}]'$. Also, recall that the conditional mean of y_t is the integral of the conditional quantiles[16] over the entire domain, that is $\mathbb{E}_{t-h}(y_t) = \int_0^1 (\alpha_0(\tau) + \alpha_1(\tau)x_{t,t-h})\, d\tau = \overline{\alpha_0} + \overline{\alpha_1}x_{t,t-h}$ or, equivalently, $x_{t,t-h} = (\mathbb{E}_{t-h}(y_t) - \overline{\alpha_0})/\overline{\alpha_1}$, where $\overline{\alpha_0} = \int_0^1 \alpha_0(\tau)d\tau$ and $\overline{\alpha_1} = \int_0^1 \alpha_1(\tau)d\tau$. This way, the optimal forecast, from Proposition 3 of Patton and Timmermann (2007), becomes:

$$\begin{aligned} \widetilde{f}_{i,t}^{h} &= \alpha_0(q_h^*) + \alpha_1(q_h^*)x_{t,t-h} \\ &= \alpha_0(q_h^*) + \alpha_1(q_h^*)(\mathbb{E}_{t-h}(y_t) - \overline{\alpha_0})/\overline{\alpha_1} \\ &= k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t), \end{aligned} \tag{20}$$

which is exactly the outcome of our Proposition 6, where we defined $k_i^h = \alpha_0(q_h^*) - (\overline{\alpha_0}/\overline{\alpha_1})\,\alpha_1(q_h^*)$ and $\beta_i^h = \alpha_1(q_h^*)/\overline{\alpha_1}$. Here, all the source of temporal dynamics is given by the covariate $x_{t,t-h}$, which simultaneously affects the conditional mean and the conditional variance.[17]

Alternative identification strategies could be pursued, e.g., a within-group method. For instance, one could explore the panel-data model proposed by Bai (2009), which allows for the joint presence of additive and interactive effects. In our notation, it would be:

$$f_{i,t}^h = X_{i,t}\gamma + \mu + k_i^h + \eta_t^h + \beta_i^h \mathbb{E}_{t-h}\left(y_t\right) + \varepsilon_{i,t}^h. \tag{21}$$

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 ($\mathbb{E}_{t-h}\left(y_t\right)$ here) requires constraining $\sum_{i=1}^N \beta_i^h = 0$ and $\sum_{t=1}^T \mathbb{E}_{t-h}\left(y_t\right) = 0$. So, although identification is possible, Bai's method imposes restrictions either on structural parameters and/or on latent factors, which our method avoids[18].

We extend the previous literature of forecasting within a panel-data framework, e.g., Palm and Zellner (1992), Davies and Lahiri (1995), Davies (2006), Issler and Lima (2009), Lahiri, Peng, and Sheng (2013), and Gaglianone and Lima (2014). Our setup encompasses that of Davies and Lahiri (and Davies), reproduced below with our

[15] See the Appendix A.3 for a discussion on the multicovariates case.

[16] See Koenker (2005, p.302).

[17] A more general setup could consider a DGP in the linear exponential family, where parameter estimates of the conditional mean can be computed even if the data density is misspecified. In addition, the conditional mean is the optimal forecast under any loss function in the Bregman class. See Gneiting (2012) and Patton (2011, 2015) for further details.

[18] Other possible route is the approach of Moreira (2009) which proposes a marginal likelihood approach to solve the incidental parameter problem based on the the maximum invariance principle. The idea is to find a group of transformations that preserves the structural parameters and yields a reduction in the incidental parameter space to a finite dimension. The methodology can be used to construct an estimator for dynamic panel data, which is consistent as $(N, T \to \infty)$ regardless of the relative rate of N and T. We leave these potential routes as suggestions for future research.

notation:

$$y_t - f_{i,t}^h = -\left(k_i^h + \eta_t^h + \varepsilon_{i,t}^h\right), \tag{22}$$

by imposing $\beta_i^h = 1$ for all $i = 1, ..., N$ and all $h = 1, ..., H$. Also, we generalize the results in Issler and Lima, where $\beta_i^h = 1$ for all $i = 1, ..., N$. Here, we have two sources of bias correction: intercept and slope. Notice that both arise from a structural affine function that links individual forecasts to the conditional expectation. In itself, this provides a general framework that can be used whenever a panel of forecasts is available.

For a given horizon h, the orthogonality conditions (13) 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, $\overline{f_{\cdot,t}^h}$ can be viewed as a *consensus forecast*, where 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. Estimation is usually done by least squares and can serve as a basis for "rationality" tests of survey expectations. It is worth noting that our setup also allows a zero-bias test within a GMM framework. If the null of zero bias is not rejected we validate the use of the consensus forecast as a consistent estimate for the conditional expectation.

On the macroeconomic literature dealing with *rational inattention* and *sticky information*, a key result is that inertia in updating information generates predictable forecast errors. Implicitly or explicitly, inertia is due to the cost of updating information from the point-of-view of the agent producing forecasts. In our context, all information is free. So, there is no cost of acquiring it. Nevertheless, we still obtain predictable forecast errors. Take equation (10) for $f_{i,t}^h$, subtract it from y_t, and take conditional expectations to obtain:

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

which is nonzero unless the restrictions $k_i^h = 0$ and $\beta_i^h = 1$ are valid for all i and h. In our setup, what generates predictable forecast errors is the interaction between an asymmetric Loss function and an unknown DGP, but not the cost of information, which is nil. So, we conclude that finding predictable forecast errors cannot be used to validate sticky information, since it also happens in models of costless information such as ours.

Still, in macroeconomics, being able to identify $\mathbb{E}_{t-h}\left(y_t\right)$ is important for identifying

economic shocks to y_t, i.e., $y_t - \mathbb{E}_{t-h}(y_t)$. Although we do not go to the length of identifying structural shocks to y_t, e.g., *supply* or *demand* shocks to y_t, we still identify shocks in a broader sense that can serve as a benchmark for alternative identifications schemes relying on specific economic models.

Finally, our method avoids dealing with sparsity in *big data* in the sense of Belloni and Chernozhukov (2011, 2013) and Belloni, Chernozhukov, and Wang (2011). The key issue is that one can think of our techniques as a combination device for survey information. Hence, agents are the ones dealing with sparse-model issues and dimension reduction techniques in producing their forecasts. We simply combine them as final users. Despite that, sparsity can arise within our context if we were unable to sample the same individuals over time in our surveys, which would generate missing data and unbalanced panels. Nonetheless, if surveys are a random sample on the cross-sectional dimension, such that all ensemble averages converge to the same limit, sparsity will pose no problem to identify and estimate our structural parameters. We leave the discussion of the more complicated case of non-random sampling for future research.

3 Empirical Application

3.1 Data

The Focus Survey of forecasts is a unique 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. Thus, 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 Carvalho and Minella (2012).

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)[19]. However, since we must rely (with respect to nowcasts) on the active

[19] The *Focus Survey* database has been collected since 1999, the year in which the Brazilian inflation-

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, $i = 1, ..., N$, inform their expectations regarding inflation rates all the way up to the next 14 months. For instance, survey respondent i informs on January 2nd, 2006 her "backcast" for the inflation rate of December, 2005 (only released on January 12th, 2006), as well as her "nowcast" for January, 2006, and the forecasts for the following 12 months, ending in January, 2007. Next working day, January 3rd, 2006, the same agent may (or may not) update the forecasts for the same inflation rates, which are treated here as events.

On January 12th, 2006 – the day in which the inflation rate of December 2005 is released – survey respondents start informing their forecasts for a different set of 14 events, now beginning in January, 2006, and ending in February, 2007. Thus, our sample covers forecasts for events (i.e. monthly IPCA inflation rates) from December 2005 to January 2014 ($T = 98$ months or events), which represents a period of stable inflation in Brazil, and the forecast horizon h ranges from one day up to $H = 400$ days. Therefore, $t = 1, ..., T$, indexes months, while $h = 1, ..., H$, indexes days.

The raw data contains forecasts for "fixed-events" and varying forecast horizons; see Bakhshi et al. (2005). To fit the setup discussed in (13) and (14), 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 dataset forms an unbalanced panel ($N \times T \times H$) containing an amount of $2,732,827$ observations. The final dataset used in this paper contains $1,486,559$ observations, since we only considered forecasts from agents that participate on the survey in a regular basis forming a balanced panel. 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 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 in *big data*, the forecasting gains we report below are non-trivial, and should be expected to increase on a truly big-data framework. So, results here 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

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.

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 (13) or (14). 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.[20]

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 four types of forecast methods. The first is the one proposed in this paper – which we have labelled the *extended bias-corrected average forecast (BCAF)*, or *extended BCAF* for short. 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 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 *extended BCAF*, we used a set of instruments containing lagged inflation π_t and interest i_t rates (IPCA and Selic, respectively). Interest rates are 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 (13) for each horizon (h), instead of using its stacked version – equation (14). In practice, estimation of the stacked version (14) was not feasible. The empirical exercise was conducted using the *R software* (version 3.0.1), 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.

[20] 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.

3.2 Empirical Results

The results of our empirical exercise are next presented. Figure 1 shows monthly IPCA inflation as well as its respective 12-month accumulated rate. Despite a spike around 2003, 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 BCAF, and the extended BCAF, at these same horizons.

Figure 1 - CPI inflation rate in Brazil (y_t)

Monthly (left plot) and twelve-month (right plot) accumulated rates (%)

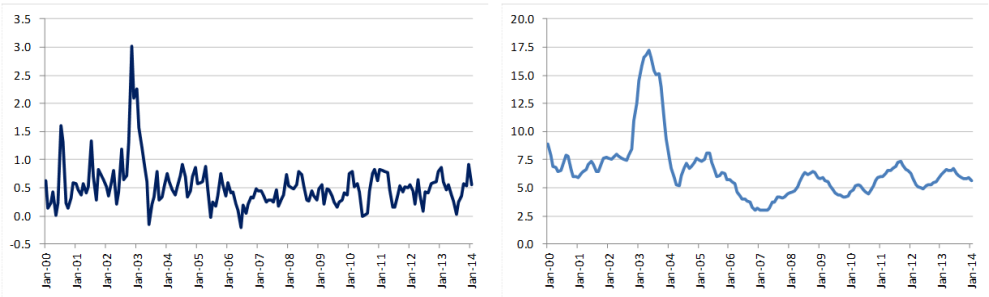
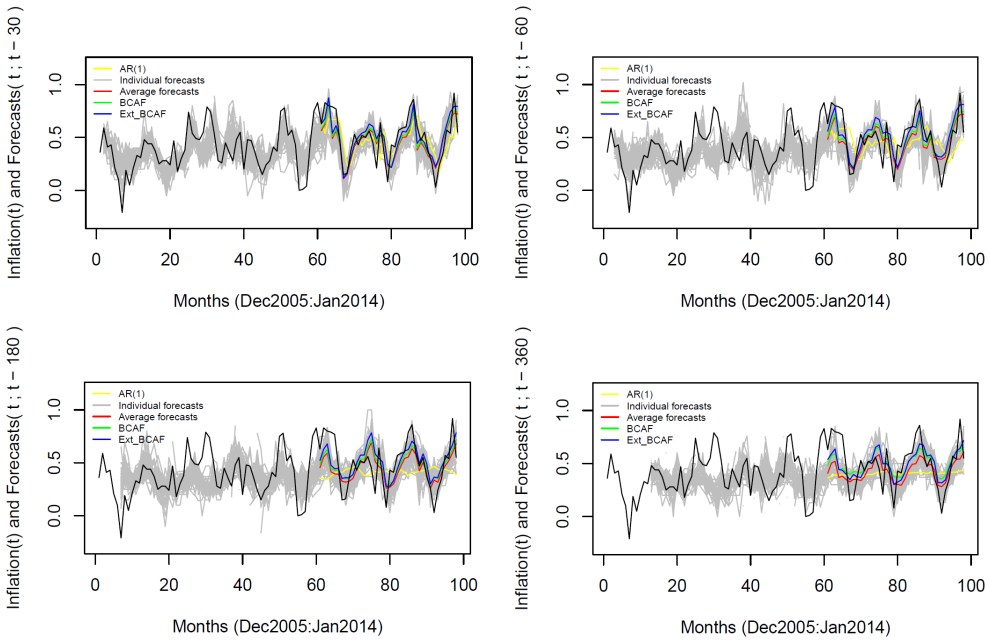


Table 1 reports GMM estimates for the extended BCAF. Although estimates of the average intercept are close to zero, and estimates of the average slope are close to unity, they are all highly significantly different to zero at any given horizon. A Wald-test for zero average intercept and unity average 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 (working days) horizons with 95% confidence bands.

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



Note: 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 lines represent the average forecast, and the green and blue lines show, respectively, the BCAF and the Extended BCAF forecasts. Black line is the inflation rate y_t.

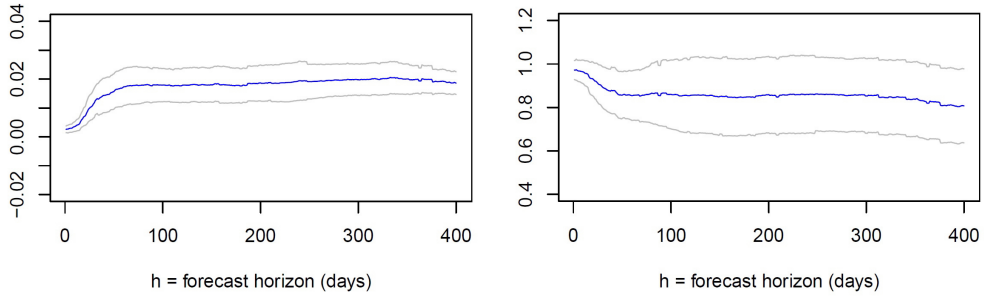
Table 1 - GMM estimation results

horizon h (days)	$\widehat{k^h}$	$\widehat{\beta^h}$	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

Note: 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 $\widehat{\overline{k^h}}$ (left panel) and $\widehat{\overline{\beta^h}}$ (right panel) and 95% confidence intervals

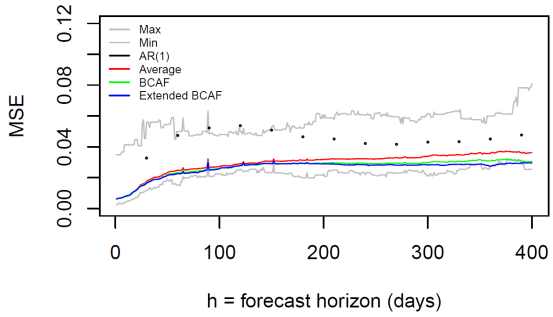


We compare the out-of-sample MSE of each of the four forecast methods employed here: the AF, the BCAF, the extended BCAF and the $AR\,(1)$ model as well as the minimum and maximum MSEs across all survey respondents included in our sample. Figure 4 plots the daily results for these methods[21]. The first interesting feature is the comparison between the $AR\,(1)$ model and the forecast-combination techniques discussed here – the AF, the BCAF, and the extended BCAF. Strikingly, notice that the $AR\,(1)$ model is usually close to upper bound of the MSE, while the first three 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).

Next, when we compare the MSE of the four methods discussed above there is an obvious *pecking order*: the extended BCAF is superior to the BCAF, which in turn dominates the AF, which in turn dominates the $AR\,(1)$ model. This validates 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, 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 for inflation can reach up to 50% for the extended BCAF.

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

Figure 4 - Mean Squared Error (MSE)



Note: Max (Min) denotes the maximum (minimum) MSE, for each horizon, across all forecasters. Average refers to the MSE of the consensus forecast.

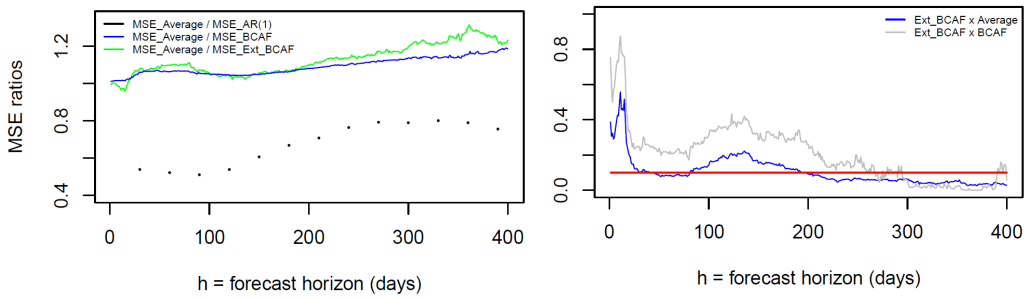
Table 2 - Mean Squared Error (MSE)

horizon h (days)	AR(1)	Average Forecast	BCAF	Extended BCAF
30	0.0328∗ [0.000]	0.0177 (0.105)	0.0166 (0.247)	0.0164
60	0.0475∗ [0.001]	0.0248∗ (0.081)	0.0234 (0.206)	0.0226
90	0.0523∗ [0.005]	0.0267 (0.119)	0.0254 (0.261)	0.0249
180	0.0466∗ [0.049]	0.0312 (0.124)	0.0294 (0.297)	0.0293
270	0.0415∗ [0.046]	0.0328∗ (0.056)	0.0293∗ (0.099)	0.0281
360	0.0452∗ [0.019]	0.0357∗ (0.028)	0.0309∗ (0.000)	0.0275

Notes: The second column shows [in brackets] the p-values of the equal variances' test of Diebold-Mariano (1995) between the AR(1) and the Extended BCAF. The third and fourth columns show (in parenthesis) the p-values of the equal-predictive accuracy 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 test for equal variances shows that the former is statistically superior to the latter (at the 10% significance level) at all horizons.

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 and Ho: equal predictive accuracy.

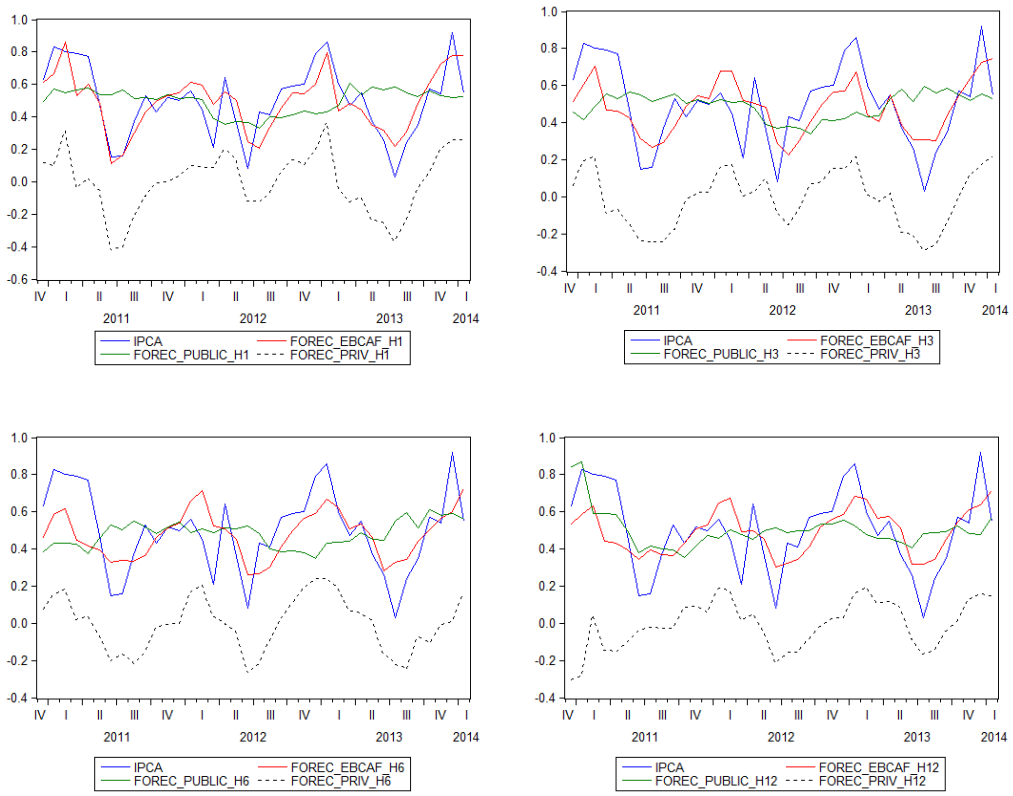
Finally, regarding heterogeneous beliefs, Proposition 11 allows us to decompose the estimated EBCAF into public and private terms, that is $\widetilde{\mathbb{E}}_{t-h}(y_t) = \mathbb{E}_{t-h}^{public}(y_t) + \mathbb{E}_{t-h}^{priv.com.}(y_t)$. Here, the public term comes from a dynamic factor model, $y_t = c + \Lambda f_t + \varepsilon_t$, where the target variable y_t is modeled as the sum of three orthogonal components: (i) the intercept c; (ii) the unobserved factors f_t; and (iii) the idiosyncratic residual ε_t. The factors f_t are derived from a set of 95 macroeconomic and financial variables (see Table A.4 in Appendix) using the Principal Component Analysis.[22]

Following the literature (e.g., Bańbura et al., 2013) we specify the factors as following a vector autoregressive (VAR) process $f_t = \Phi(L)f_t + v_t$. By estimating the OLS regression $y_t = c + \Lambda f_t + \varepsilon_t$, we recover the public information term $\widehat{\mathbb{E}}_{t-h}^{public}(y_t) = \widehat{c} + \widehat{\Lambda} f_{t,t-h}$, where $f_{t,t-h}$ are $h-$step ahead out-of-sample forecasts from the VAR model estimated in a recursive scheme. The private information term is, thus, computed by $\widehat{\mathbb{E}}_{t-h}^{priv.com.}(y_t) = \widehat{\widetilde{\mathbb{E}}}_{t-h}(y_t) - \widehat{\mathbb{E}}_{t-h}^{public}(y_t)$. Figure 6 presents the results for selected horizons h.

Overall, the public information series behave like a core inflation-type series, when compared to the target variable y_t, whereas the private information series seems to embody some seasonal dynamics. In addition, the private information series $\widehat{\mathbb{E}}_{t-h}^{priv.com.}(y_t)$, for $h = 1, 3, 6$ and 12, does not Granger cause, at 5% level of significance, the interest rate (Selic), the stock market index (Ibovespa) variation in twelve months or the country risk-premium (Embi+Br). On the other hand, there is evidence that the Embi+Br series Granger-cause, at a 10% level, the private series for $h = 1$, indicating that risk-premium variations anticipate the private information part of the short-term inflation expectations (e.g., by incorporating into the nowcasts some possible inflationary shock).

[22] From the set of 95 variables, we extract the first 6 principal components (which jointly accounts for 75% of the total variance in the data) to form a monthly VAR with 2 lags (following the Hannan-Quinn information criterion).

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



4 Conclusion

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 – *big data*. Our focus here is on the abundant information contained in expectation surveys that can now be built on thousands (or millions) of respondents on an almost continuous-time basis, and how it can be used efficiently for forecasting purposes. We think of this problem as having two layers. In the first layer, there is the decision of survey respondents on what forecast should be supplied, considering a given target variable of the survey. On the second layer, there is a user of the survey who wants to forecast efficiently the target variable using all the survey information. We think of the latter as the *econometrician*.

For the decision in the first layer, we employ standard microeconomic and econometric techniques to model the optimal survey response. Under suitable conditions that are standard in the literature, we show that the optimal forecast of a given survey respondent is an affine function of the conditional expectation of the target variable of the survey. Assuming that the final user of that information has a mean-squared-error (MSE) *risk function*, we show how to estimate the optimal forecast (conditional ex-

pectation) from the point-of-view of the econometrician. Our solution can be viewed as a solution to a signal-extraction problem, where potentially the number of respondents and of time-series observations is large, i.e., N and T grow without bound. We show how to solve this problem using Hansen's (1982) generalized method of moments (GMM) in the sequential asymptotic context of Phillips and Moon (1999). We also solve the simpler problem where N is fixed, which applies to current macroeconomic surveys where there is a limited number of surveys respondents, e.g., the *Livingston* Survey of the Philadelphia FED or the *Survey of Professional Forecasters*.

Our solution can be viewed as forecast-combination technique for large N and T (or large T and fixed N). Since individual forecasts have an intercept and a slope bias vis-à-vis the conditional expectation, the optimal forecast of the second layer of decision making must correct survey forecasts accordingly. We show how to avoid the curse of dimensionality in implementing these bias corrections using the aggregation techniques proposed in Driscoll and Kraay (1998).

We apply the techniques advanced in this paper to forecast Brazilian CPI inflation using the *Focus Survey*, constructed and kept by the Central Bank of Brazil. It is a unique panel database, which could potentially collect 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 dataset used in this paper contains $1,486,559$ observations, since we only considered the forecasts from agents that participate on the survey in a regular basis, i.e., forming a balanced panel[23].

We compare the truly out-of-sample MSE of four different forecast methods – three of them widely used in this literature: the method proposed here – the *extended* bias-corrected average forecast (BCAF), the average forecast, the BCAF, and the forecast of the $AR(1)$ model for inflation. There is *pecking order*: the extended BCAF is superior to the BCAF, which in turn dominates the AF, 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%.

Possible extensions of this paper include investigating the dynamics of the forecast revision process, i.e., of $f_{i,t}^h - f_{i,t}^{h+1}$ and/or $f_{i,t}^h - f_{i,t+1}^h$, and to pursue alternative identification and estimation strategies than the ones proposed here.

[23] Dealing with unbalanced panel issues could increase the sample size to $2,732,827$ observations.

References

[1] Andrade, P., Le Bihan, H., 2013. Inattentive professional forecasters. *Journal of Monetary Economics* 60 (8), 967-982.

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

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

[4] 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.

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

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

[7] 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.

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

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

[10] Belloni, A., Chernozhukov, V., 2011. L1-penalized quantile regression for high dimensional sparse models. *Annals of Statistics* 39 (1), 82–130.

[11] Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19 (2), 521–547.

[12] Belloni, A., Chernozhukov, V., Wang, L., 2011. Square-Root Lasso: Pivotal Recovery of Sparse Signals via Conic Programming. *Biometrika* 98 (4), 791-806.

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

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

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

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

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

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

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

[20] 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.

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

[22] 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

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

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

[25] Einav, L., Levin, J.D., 2013. The Data Revolution and Economic Analysis. NBER Working Paper No. 19035.

[26] 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.

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

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

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

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

[31] 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.

[32] 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.

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

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

[35] 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.

[36] Galvao, A.F., Wang, L., 2015. Efficient minimum distance estimator for quantile regression fixed effects panel data. *Journal of Multivariate Analysis* 133, 1-26.

[37] Gneiting, T., 2012. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* 106 (494), 746-762.

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

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

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

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

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] Joseph, A., Vodenska, I., Stanley, E., Chen, G., 2014. Netconomics: Novel Forecasting Techniques from the Combination of Big Data, Network Science and Economics. Manuscript. Available at http://arxiv.org/pdf/1403.0848v1.pdf

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

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

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

[50] Lahiri, K., Peng, H., Sheng, X., 2013. Measuring Uncertainty of a Combined Forecast and a New Test for Forecaster Homogeneity. Mimeo: University at Albany, SUNY.

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

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

[53] 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.

[54] Moreira, M.J., 2009. A maximum likelihood method for the incidental parameter problem. *The Annals of Statistics* 37 (6A), 3660-3696.

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

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

[57] Patton, A.J., 2011. Volatility Forecast Comparison using Imperfect Volatility Proxies. *Journal of Econometrics* 160 (1), 246-256.

[58] Patton, A.J., 2015. Comparing Possibly Misspecified Forecasts. Working Paper, Duke University.

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

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

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

[62] 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.

[63] 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.

[64] 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.

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

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

[67] 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.

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

[69] Warin, T., Sanger, W., 2014. Structuring Big Data: How Financial Models May Help. *Journal of Computer Science and Technology* 2 (1), 1-20.

A Appendix

A.1 Proofs of Propositions

Proof of Proposition 1. Define the objective function $\Phi(\widetilde{y}) \equiv \mathbb{E}\left(L(y-\widetilde{y}) \mid \mathcal{F}_{t-h}\right) = \int_{-\infty}^{\infty} L(y-\widetilde{y})f(y)dy = \int_{-\infty}^{\widetilde{y}} \beta_1 g(y-\widetilde{y})f(y)dy + \int_{\widetilde{y}}^{\infty} \beta_2 g(y-\widetilde{y})f(y)dy$, where the last equality is due to A3, and recall that $e = y-\widetilde{y}$. Differentiating $\Phi(\widetilde{y})$ with respect to $\widetilde{y}$ and equating it to zero gives us the optimal forecast $\widetilde{y}^*$. Thus, $\partial\Phi(\widetilde{y})/\partial\widetilde{y} = 0 \therefore \beta_1 \int_{-\infty}^{\widetilde{y}^*} g'(e)f(y)dy = -\beta_2 \int_{\widetilde{y}^*}^{\infty} g'(e)f(y)dy$ or $\beta_1\gamma_1 = -\beta_2\gamma_2$, where $\gamma_1 \equiv \int_{-\infty}^{\widetilde{y}^*} g'(e)f(y)dy \geq 0$, since $f(y) \geq 0$, for all $y \in \mathbb{R}$, and from A2-A3, $g'(e) = \partial g/\partial\widetilde{y} = (\partial g/\partial e)(\partial e/\partial\widetilde{y}) = -\partial g/\partial e \geq 0$ for all $e \leq 0$ or $y \in (-\infty; \widetilde{y}^*]$. In addition, define $\gamma_2 \equiv \int_{\widetilde{y}^*}^{\infty} g'(e)f(y)dy \leq 0$, since $g'(e) \leq 0$ for all $e \geq 0$, or $y \in [\widetilde{y}^*; \infty)$, from A2-A3.

(i) If $\beta_1 > \beta_2$, then $\gamma_1 < -\gamma_2$, since $\beta_1\gamma_1 = -\beta_2\gamma_2$, $\{\beta_1; \beta_2\} > 0$, $\gamma_1 \geq 0$ and $\gamma_2 \leq 0$. This way, $\int_{-\infty}^{\widetilde{y}^*} g'(e)f(y)dy < -\int_{\widetilde{y}^*}^{\infty} g'(e)f(y)dy$. Since $g(e)$ is symmetric around $e = 0$, that is $g(e) = g(-e)$, and g' exists almost everywhere, then g' is antisymmetric around $e = 0$, so that $g'(e) = -g'(-e)$. Thus, one can rewrite the previous inequality as $\int_{-\infty}^{\widetilde{y}^*} g_1(e)f(y)dy < \int_{\widetilde{y}^*}^{\infty} g_1(e)f(y)dy$, where $g_1(e) \equiv \begin{cases} g'(e) & ; \ e \leq 0 \\ -g'(e) & ; \ e > 0 \end{cases}$. Since $g'(e) \geq 0$ for all $e \leq 0$, and $-g'(e) \geq 0$ for all $e > 0$, then $g_1(e) \geq 0$, for all $e \in \mathbb{R}$, and symmetric around $e = 0$. By using the *Second Mean Value Theorem for Integrals* on each side of the previous inequality of integrals, it follows that $g_1(\xi_L) \int_{-\infty}^{\widetilde{y}^*} f(y)dy < g_1(\xi_U) \int_{\widetilde{y}^*}^{\infty} f(y)dy$, for some $\xi_L \in (-\infty; \widetilde{y}^*]$ and some $\xi_U \in [\widetilde{y}^*; \infty)$. Since $\mathbb{E}\left(|g'(e)|\right) = \mathbb{E}\left(g_1(e)\right) \leq \int_{-\infty}^{\widetilde{y}^*} g_1(e)f(y)dy / \int_{-\infty}^{\widetilde{y}^*} f(y)dy \therefore \left(\int_{-\infty}^{\widetilde{y}^*} g_1(e)f(y)dy + \int_{\widetilde{y}^*}^{\infty} g_1(e)f(y)dy\right) \int_{-\infty}^{\widetilde{y}^*} f(y)dy \leq \int_{-\infty}^{\widetilde{y}^*} g_1(e)f(y)dy \therefore \int_{\widetilde{y}^*}^{\infty} g_1(e)f(y)dy \int_{-\infty}^{\widetilde{y}^*} f(y)dy \leq \int_{-\infty}^{\widetilde{y}^*} g_1(e)f(y)dy \left(1 - \int_{-\infty}^{\widetilde{y}^*} f(y)dy\right) \therefore \int_{\widetilde{y}^*}^{\infty} g_1(e)f(y)dy \Big/ \int_{\widetilde{y}^*}^{\infty} f(y)dy \leq \int_{-\infty}^{\widetilde{y}^*} g_1(e)f(y)dy \Big/ \int_{-\infty}^{\widetilde{y}^*} f(y)dy \therefore g_1(\xi_U) \leq g_1(\xi_L)$. This way,

$g_1(\xi_L)\int_{-\infty}^{\widetilde{y}^*} f(y)dy < g_1(\xi_U)\int_{\widetilde{y}^*}^{\infty} f(y)dy \leq g_1(\xi_L)\int_{\widetilde{y}^*}^{\infty} f(y)dy$. Since $g_1(e) \geq 0$ for all $e \in \mathbb{R}$, it follows that $\int_{-\infty}^{\widetilde{y}^*} f(y)dy < \int_{\widetilde{y}^*}^{\infty} f(y)dy = 1 - \int_{-\infty}^{\widetilde{y}^*} f(y)dy$. Thus, $\int_{-\infty}^{\widetilde{y}^*} f(y)dy < 0.5$ $\therefore \widetilde{f}^h_{i,t} < Med_{t-h}(y_t) < \mathbb{E}_{t-h}(y_t)$.

On the other hand, since $\beta_1\gamma_1 = \beta_2(-\gamma_2) \geq 0$ and $\beta_1^i/\beta_2^i > \beta_1^j/\beta_2^j$, then $-\gamma_2^i/\gamma_1^i > -\gamma_2^j/\gamma_1^j \therefore -\gamma_2^i\gamma_1^j > -\gamma_2^j\gamma_1^i \therefore \int_{\widetilde{y}_i^*}^{\infty} -g'(e_i)f(y)dy \int_{-\infty}^{\widetilde{y}_j^*} g'(e_j)f(y)dy > \int_{\widetilde{y}_j^*}^{\infty} -g'(e_j)f(y)dy \int_{-\infty}^{\widetilde{y}_i^*} g'(e_i)f(y)dy$. Thus, $\int_{\widetilde{y}_i^*}^{\infty} g_1(e_i)f(y)dy \int_{-\infty}^{\widetilde{y}_j^*} g_1(e_j)f(y)dy > \int_{\widetilde{y}_j^*}^{\infty} g_1(e_j)f(y)dy \int_{-\infty}^{\widetilde{y}_i^*} g_1(e_i)f(y)dy$. Since $g_1(e_j)$ is symmetric around $e_j = 0$, it follows that $\int_{\widetilde{y}_i^*}^{\infty} g_1(e_i)f(y)dy \int_{-\infty}^{\widetilde{y}_j^*} f(y)dy > \int_{\widetilde{y}_j^*}^{\infty} f(y)dy \int_{-\infty}^{\widetilde{y}_i^*} g_1(e_i)f(y)dy$. Now, since $g_1(e_i)$ is symmetric around $e_i = 0$, it follows that $\int_{\widetilde{y}_i^*}^{\infty} f(y)dy \int_{-\infty}^{\widetilde{y}_j^*} f(y)dy > \int_{\widetilde{y}_j^*}^{\infty} f(y)dy \int_{-\infty}^{\widetilde{y}_i^*} f(y)dy \therefore \left(1 - \int_{-\infty}^{\widetilde{y}_i^*} f(y)dy\right) \int_{-\infty}^{\widetilde{y}_j^*} f(y)dy > \left(1 - \int_{-\infty}^{\widetilde{y}_j^*} f(y)dy\right) \int_{-\infty}^{\widetilde{y}_i^*} f(y)dy \therefore \int_{-\infty}^{\widetilde{y}_j^*} f(y)dy > \int_{-\infty}^{\widetilde{y}_i^*} f(y)dy \therefore F_{t,t-h}(\widetilde{y}_j^*) > F_{t,t-h}(\widetilde{y}_i^*)$ $\therefore \widetilde{f}^h_{j,t} > \widetilde{f}^h_{i,t}$. Since $\beta_1^j/\beta_2^j > 1 \therefore \widetilde{f}^h_{j,t} < \mathbb{E}_{t-h}(y_t)$, thus $\widetilde{f}^h_{i,t} < \widetilde{f}^h_{j,t} < \mathbb{E}_{t-h}(y_t)$.

(ii) If $\beta_1 < \beta_2$, based on similar arguments of the proof for item (i), it follows that $\widetilde{f}^h_{i,t} > Med_{t-h}(y_t) > \mathbb{E}_{t-h}(y_t)$ and $\widetilde{f}^h_{i,t} > \widetilde{f}^h_{j,t} > \mathbb{E}_{t-h}(y_t)$. ■

Proof of Proposition 2. (i) If $\beta_1 > \beta_2$, then from the proof of Proposition 1, it follows that $\int_{-\infty}^{\widetilde{y}^*} g_1(e)f(y)dy < \int_{\widetilde{y}^*}^{\infty} g_1(e)f(y)dy$. Since f is a unimodal and symmetric function around $\mathbb{E}_{t-h}(y_t) = Med_{t-h}(y_t)$, and from A2-A3 g_1 is symmetric around $e = 0$ or $y = \widetilde{y}^*$, and by using the fact that both f and g_1 are non-negative for all $y \in \mathbb{R}$, it follows that $\widetilde{f}^h_{i,t} < \mathbb{E}_{t-h}(y_t)$.

(ii) If $\beta_1 < \beta_2$, then based on a similar argument, it follows that $\widetilde{f}^h_{i,t} > \mathbb{E}_{t-h}(y_t)$. ■

Proof of Proposition 3. See Granger (1969). ■

Proof of Proposition 4. See Granger and Newbold (1986, p.126). ■

Proof of Proposition 5. In the proof of Proposition 1, the optimal forecast $\widetilde{f}^h_{i,t}$ is viewed as a conditional quantile of y_t, such that $\tau_i \equiv F_{t,t-h}(\widetilde{f}^h_{i,t})$.

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

(ib) Two-piece normal $(\mu, \sigma_1^2, \sigma_2^2)$: according to Julio (2007), the conditional quantiles of a two-piece normal distribution, where μ denotes[24] $\mathbb{E}_{t-h}(y_t)$, are given by: $\mu - \sqrt{2/\pi}(\sigma_2 - \sigma_1) + \sigma_1\Phi^{-1}(\frac{\tau}{C\sqrt{2\pi}\sigma_1})$ where $C = \sqrt{\frac{2}{\pi}}(\sigma_1 + \sigma_2)^{-1}$ for $\tau \leq \Pr[x \leq$

[24] Julio (2007) uses μ to denote the mode, whereas in this proof we use μ to denote the mean = mode + $\sqrt{2/\pi}(\sigma_2 - \sigma_1)$.

$\mu - \sqrt{2/\pi}(\sigma_2 - \sigma_1)]$; or equal to $\mu - \sqrt{2/\pi}(\sigma_2 - \sigma_1) + \sigma_2\Phi^{-1}(\frac{\tau + C\sqrt{2\pi}\sigma_2 - 1}{C\sqrt{2\pi}\sigma_2})$ for $\tau > \Pr[x \leq \mu - \sqrt{2/\pi}(\sigma_2 - \sigma_1)]$. In both cases, notice that $\widetilde{f}^h_{i,t} = \mathbb{E}_{t-h}(y_t) + k^h_i$.

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

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

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

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

(iv) Beta $(a > 0, b = 1)$: The conditional quantiles are given by $\tau^{1/a}$, then, $\ln(\widetilde{f}^h_{i,t}) = \ln(\tau_i^{1/a}) \therefore a = \frac{\ln(\tau_i)}{\ln(\widetilde{f}^h_{i,t})}$. Since $\mathbb{E}_{t-h}(y_t) = \frac{a}{a+b} = \frac{a}{a+1} = \frac{\frac{\ln(\tau_i)}{\ln(\widetilde{f}^h_{i,t})}}{1+\frac{\ln(\tau_i)}{\ln(\widetilde{f}^h_{i,t})}} = \frac{\ln(\tau_i)}{\ln(\tau_i)+\ln(\widetilde{f}^h_{i,t})} \therefore \ln(\widetilde{f}^h_{i,t}) = \frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)} - \ln(\tau_i)$

$\therefore \widetilde{f}^h_{i,t} = \exp(\frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)})\exp(-\ln(\tau_i))$. Therefore, $\widetilde{f}^h_{i,t} = \beta^h_i\varphi(\mathbb{E}_{t-h}(y_t))$, in which $\beta^h_i = \exp(-\ln(\tau_i))$ and φ is the non-linear function $\varphi(\mathbb{E}_{t-h}(y_t)) = \exp(\frac{\ln(\tau_i)}{\mathbb{E}_{t-h}(y_t)})$. ■

Proof of Proposition 6. (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:

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

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 $\widetilde{f}_{i,t}^{h}$ is given by:

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

in which we have used the fact that $F_{\eta,h}(\eta)$ is time-invariant by definition, and $\gamma_h^i \equiv \underset{\widetilde{\gamma}}{\arg\min} \int L^i (\eta_t - \widetilde{\gamma}) \, dF_{\eta,h}$ or, equivalently, $\gamma_h^i = F_{\eta,h}^{-1}(\tau_i)$ since $\widetilde{f}_{i,t}^{h} = F_{t,t-h}^{-1}(\tau_i)$. 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 $\widetilde{f}_{i,t}^{h} = \alpha_0(\tau_i) + \alpha_1(\tau_i) x_{t,t-h}$.

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 $F_{t,t-h}^{-1}(\tau) = \alpha_0(\tau) + \alpha_1(\tau) x_{t,t-h}$, for some $\tau \in [0,1] \therefore \mathbb{E}_{t-h}(y_t) = \int_0^1 F_{t,t-h}^{-1}(\tau) d\tau = \int_0^1 (\alpha_0(\tau) + \alpha_1(\tau) x_{t,t-h}) \, d\tau = \overline{\alpha_0} + \overline{\alpha_1} x_{t,t-h}$, where $\overline{\alpha_j} = \int_0^1 \alpha_j(\tau) d\tau$; $j = \{0;1\}$. Thus, $\mathbb{E}_{t-h}(y_t) = \overline{\alpha_0} + \overline{\alpha_1} x_{t,t-h} \therefore \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} \mathbb{E}_{t-h}(y_t) = \overline{\alpha_0} \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} + \overline{\alpha_1} \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} x_{t,t-h} \therefore \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} \mathbb{E}_{t-h}(y_t) + \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}} \alpha_1(\tau_i) \right) = \overline{\alpha_0} \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} + \overline{\alpha_1} \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}} x_{t,t-h}$ $+ \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}} \alpha_1(\tau_i) \right) = \alpha_0(\tau_i) + \alpha_1(\tau_i) x_{t,t-h} = \widetilde{f}_{i,t}^{h}$. Therefore, if one defines $k_i^h \equiv \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}} \alpha_1(\tau_i) \right)$ and $\beta_i^h \equiv \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}$ it follows that $\widetilde{f}_{i,t}^{h} = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t)$. Notice that k_i^h and β_i^h 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_h^i, which is a constant that depends only on the distribution $F_{\eta,h}(0,1)$ and 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 $\tau \in [0,1] \therefore \beta_i^h = 1$. ■

Proof of Proposition 7. From Proposition 6, it follows that $\widetilde{f}_{i,t}^{h} = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t)$, where $k_i^h \equiv \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}} \alpha_1(\tau_i) \right)$ and $\beta_i^h \equiv \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}$; $\alpha_0(\tau_i) = (\delta_0 + \gamma_0 \gamma_h^i)$; $\alpha_1(\tau_i) = (\delta_1 + \gamma_1 \gamma_h^i)$; $\overline{\alpha_j} = \int_0^1 \alpha_j(\tau) d\tau$ 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_i^h and β_i^h 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_h^i, 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 $f_{i,t}^{h}$ of y_t conditioned on the information set available at period $(t-h)$ is given by $f_{i,t}^{h} = \widehat{k_i^h} + \widehat{\beta_i^h} \widehat{\mathbb{E}}_{t-h}(y_t)$. If we define the error term $\varepsilon_{i,t}^{h}$ as the difference between the optimal (feasible) forecast and the optimal forecast, it follows that $\varepsilon_{i,t}^{h} \equiv f_{i,t}^{h} - \widetilde{f}_{i,t}^{h} = f_{i,t}^{h} - k_i^h - \beta_i^h \mathbb{E}_{t-h}(y_t) \therefore f_{i,t}^{h} = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^{h}$, where $\varepsilon_{i,t}^{h} \neq 0$, provided that $\widehat{\alpha}_0(\tau) - \alpha_0(\tau) \neq 0$ and

$\widehat{\alpha}_1(\tau) - \alpha_1(\tau) \neq 0$, for all $\tau \in [0;1]$, due to parameter uncertainty, under a finite sample with T observations of $\{y_t; x_{t,t-h}\}_{t=1}^{T}$. Now, in order to prove consistency of $[\widehat{k_i^h}; \widehat{\beta_i^h}]$, define $\widehat{\alpha}(\tau) \equiv [\widehat{\alpha}_0(\tau); \widehat{\alpha}_1(\tau)]$. Koenker (2005, Theorem 4.1, p.120) shows that the estimator $\widehat{\alpha}_n(\tau) = \underset{\alpha \in \mathbb{R}^2}{\arg\min} \sum_{t=1}^{n} \rho_\tau(y_t - \alpha_0 - \alpha_1 x_{t,t-h})$, where ρ_τ is defined as in Koenker and Basset (1978) by $\rho_\tau(u) = \begin{cases} \tau u, u \geq 0 \\ (\tau - 1)u, u < 0 \end{cases}$, is a consistent estimator for $\alpha(\tau)$, in linear conditional quantile functions as long as $n \to \infty$ (under conditions A1-A2 of Koenker (2005, p.120)). This way, it follows that $\widehat{\widehat{\alpha}} = \sum_{k=1}^{K} \widehat{\alpha}(\tau_k)\Delta\tau_k \xrightarrow{p} \sum_{k=1}^{K} \alpha(\tau_k)\Delta\tau_k$, 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 $\lim_{K\to\infty}\left(\sum_{k=1}^{K} \alpha(\tau_k)\Delta\tau_k\right) = \int_0^1 \alpha(\tau)d\tau = \overline{\alpha}$, 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 $\widehat{k_i^h} \equiv \left(\widehat{\alpha}_0(\tau_i) - \frac{\widehat{\overline{\alpha_0}}}{\widehat{\overline{\alpha_1}}}\widehat{\alpha}_1(\tau_i)\right) \xrightarrow{p} k_i^h = \left(\alpha_0(\tau_i) - \frac{\overline{\alpha_0}}{\overline{\alpha_1}}\alpha_1(\tau_i)\right)$ and $\widehat{\beta_i^h} \equiv \frac{\widehat{\alpha}_1(\tau_i)}{\widehat{\overline{\alpha_1}}} \xrightarrow{p} \beta_i^h = \frac{\alpha_1(\tau_i)}{\overline{\alpha_1}}$. ■

Proof of Proposition 8. Assumptions A1-A6 and Proposition 7 are necessary to justify an optimal feasible forecast of y_t conditioned on $\mathcal{F}_{t-h}$ of the form $f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h$. Then, assumptions A7-A9 are used to achieve the set of orthogonality conditions (13) or (14), 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 $\operatorname{plim}_{(N,T\to\infty)_{\text{seq}}} \left(\frac{f_{\cdot,t}^h - \widehat{k^h}}{\widehat{\beta^h}}\right) = y_t + \operatorname{plim}_{T\to\infty} \left(\frac{\operatorname{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \widehat{v_{i,t}^h}}{\widehat{\beta^h}}\right)$

$= y_t + \frac{\underset{(N,T\to\infty)_{\text{seq}}}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\widehat{v_{i,t}^h}}{\underset{T\to\infty}{\text{plim}}\left(\widehat{\beta^h}\right)} = y_t + \frac{\beta^h \eta_t^h}{\beta^h} = \mathbb{E}_{t-h}(y_t)$, since (from A8-A9) it follows that $\underset{T\to\infty}{\text{plim}}\left(\widehat{\beta^h}\right) = \beta^h = \beta_0^h \neq 0$, 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 (13) or (14) for each h separately. ■

Proof of Proposition 9. 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, $\mathbb{E}_{t-h}(y_t) = \left(\frac{f_{i,t}^h - k_i^h}{\beta_i^h}\right)$, for all i, t and given h. It also follows that $\mathbb{E}_{t-h}(y_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - k_i^h}{\beta_i^h} = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}}$, where the first equality is simply the cross-section average applied to the equation $\mathbb{E}_{t-h}(y_t) = \left(\frac{f_{i,t}^h - k_i^h}{\beta_i^h}\right)$, 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: $\widehat{\mathbb{E}}_{t-h}(y_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}$. Then, let $T \to \infty$ to obtain $\underset{T\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}} = \mathbb{E}_{t-h}(y_t)$, due to A11 (consistency of $\widehat{\overline{\theta^h}} = \left[\widehat{\overline{k^h}}; \widehat{\overline{\beta^h}}\right]'$ estimates for all N). In addition, letting (sequentially) $N \to \infty$ does not change the convergence to the conditional mean, that is, $\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}} = \mathbb{E}_{t-h}(y_t)$, 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 (13) or (14) for each h separately. ■

Proof of Proposition 10. Assumptions A1-A11 provide sufficient conditions for the sequential convergences $(N, T \to \infty)_{\text{seq}}$ and $(T, N \to \infty)_{\text{seq}}$ to hold in our setup (see Propositions 8 and 9, respectively). By also assuming conditions A12 and A13, joint convergence directly follows from Lemma 6a of Phillips and Moon (1999). ■

Proof of Proposition 11. Following the proof of Proposition 7, 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 8, it follows (from A8 and A15) 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 A14-A15, 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: $y_t = \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{i,t-h}(y_t)\right) - \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_{i,t}^h\right) = \widetilde{\mathbb{E}}_{t-h}(y_t) - \eta_t^h$. On the other hand, from A16 it follows that $\mathcal{F}_{t-h} \cap \mathcal{F}_{i,t-h}^{PC} = \emptyset$, $\mathcal{F}_{t-h} \cap \mathcal{F}_{i,t-h}^{PI} = \emptyset$, $\mathcal{F}_{i,t-h}^{PC} \cap \mathcal{F}_{i,t-h}^{PI} = \emptyset$ and $\mathcal{F}_{i,t-h} = \mathcal{F}_{t-h} \cup \mathcal{F}_{i,t-h}^{PC} \cup \mathcal{F}_{i,t-h}^{PI} \neq \emptyset$, i.e., $\mathcal{F}_{t-h}$, $\mathcal{F}_{i,t-h}^{PC}$ and $\mathcal{F}_{i,t-h}^{PI}$ form a *partition* of $\mathcal{F}_{i,t-h}$. Therefore, from the properties of conditional expectation as an orthogonal projection operator, it follows that $\mathbb{E}_{i,t-h}(y_t) = \mathbb{E}[y_t|\mathcal{F}_{t-h}] + \mathbb{E}\left[y_t\middle|\mathcal{F}_{i,t-h}^{PC}\right] + \mathbb{E}\left[y_t\middle|\mathcal{F}_{i,t-h}^{PI}\right]$. By taking the cross-sectional average on both sides of the previous expression, and letting $N \to \infty$, it follows that: $\widetilde{\mathbb{E}}_{t-h}(y_t) = \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(y_t \mid \mathcal{F}_{t-h})\right) + \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{PC}\right)\right)$, since from A17: $\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{PI}\right)\right) = 0$. Therefore, from A18, it follows that: $\widetilde{\mathbb{E}}_{t-h}(y_t) = \mathbb{E}_{t-h}^{public}(y_t) + \mathbb{E}_{t-h}^{priv.com.}(y_t)$. ■

A.2 Parameter uncertainty

Example 1 - Suppose that:

(i) DGP is given by $\left(y_{t|t-h}\right) \sim N(\mu_{t|t-h}, \sigma^2)$;

(ii) Loss function of individual i is the MSE function;

(iii) Individual i (correctly) assumes that the DGP is given by a Gaussian distribution with conditional mean dynamics, however, due to finite sample, she/he estimates $(\widehat{\mu}_{t|t-h}, \widehat{\sigma^2})$.

In this case, from (i) and (ii), the optimal forecast (Granger, 1969) is $\widetilde{f}_{i,t}^h = \mathbb{E}_{t-h}(y_t) = \mu_{t|t-h}$. However, the best that individual i can do is to approximate the "optimal forecast" by using (ii) and (iii), so that $f_{i,t}^h = \widehat{\mathbb{E}}_{t-h}(y_t) = \widehat{\mu}_{t|t-h}$. Thus, the

optimal forecast "approximation error" is given by $\varepsilon_{i,t}^{h} \equiv f_{i,t}^{h} - \widetilde{f}_{i,t}^{h} = \widehat{\mu}_{t|t-h} - \mu_{t|t-h}$. Provided that $\widehat{\mu}_{t|t-h}$ is a consistent estimator for $\mu_{t|t-h}$, it follows that $\underset{T\to\infty}{\text{plim}}\left(\varepsilon_{i,t}^{h}\right) = 0$.

Example 2 - Suppose that:
(i) DGP is given by $\left(y_{t|t-h}\right) \sim N(\mu_{t|t-h}, \sigma^2)$;
(ii) Loss of individual i is the Lin-Lin asymmetric function of Proposition 4;
(iii) Individual i (correctly) assumes that the DGP is given by a Gaussian distribution with conditional mean dynamics, however, due to finite sample, she/he estimates $(\widehat{\mu}_{t|t-h}, \widehat{\sigma^2})$.

In this case, from (i) and (ii), the optimal forecast (see Propositions 4-5) is $\widetilde{f}_{i,t}^{h} = k_i^h + \mathbb{E}_{t-h}(y_t)$, where $k_i^h = \sigma\Phi^{-1}(\tau_i)$ and $\tau_i = \beta_2/(\beta_1+\beta_2)$. However, individual i approximates the "optimal forecast" by using (ii) and (iii), so that $f_{i,t}^{h} = \widehat{\mu}_{t|t-h} + \widehat{\sigma}\Phi^{-1}(\tau_i)$. Thus, the optimal forecast "approximation error" is given by $\varepsilon_{i,t}^{h} \equiv f_{i,t}^{h} - \widetilde{f}_{i,t}^{h} = \left(\widehat{\mu}_{t|t-h} - \mu_{t|t-h}\right) + \left(\widehat{\sigma} - \sigma\right)\Phi^{-1}(\tau_i)$. Provided that $(\widehat{\mu}_{t|t-h}, \widehat{\sigma^2})$ are consistent estimators for $(\mu_{t|t-h}, \sigma^2)$, it follows that $\underset{T\to\infty}{\text{plim}}\left(\varepsilon_{i,t}^{h}\right) = 0$.

Example 3 - Suppose that:
(i) DGP is given by $\left(y_{t|t-h}\right) \sim Weibull(b,k)$;
(ii) Loss of individual i is the Lin-Lin asymmetric function of Proposition 4;
(iii) Individual i (wrongly) assumes that the DGP is given by a Gaussian distribution with conditional mean dynamics and, thus, she/he estimates $(\widehat{\mu}_{t|t-h}, \widehat{\sigma^2})$.

In this case, from (i) and (ii), the optimal forecast (see proof of Proposition 5) is $\widetilde{f}_{i,t}^{h} = b\left[-\ln(1-\tau_i)\right]^{1/k}$, where $\tau_i = \beta_2/(\beta_1+\beta_2)$. However, individual i try to approximate the "optimal forecast" by using (ii) and (iii), so that $f_{i,t}^{h} = \widehat{\mu}_{t|t-h} + \widehat{\sigma}\Phi^{-1}(\tau_i)$. This way, $\varepsilon_{i,t}^{h} \equiv f_{i,t}^{h} - \widetilde{f}_{i,t}^{h} = \widehat{\mu}_{t|t-h} + \widehat{\sigma}\Phi^{-1}(\tau_i) - b\left[-\ln(1-\beta_2/(\beta_1+\beta_2))\right]^{1/k}$ and, thus, $\underset{T\to\infty}{\text{plim}}\left(\varepsilon_{i,t}^{h}\right) \neq 0$.

A.3 Location-scale model with k covariates

The linear conditional quantile of y_t assuming a location-scale model with k covariates is given by $F_{t,t-h}^{-1}(\tau) = \alpha_0(\tau) + \alpha_1(\tau)x_{1,t-h} + ... + \alpha_k(\tau)x_{k,t-h} = \alpha_0(\tau) + \sum_{j=1}^{k}\alpha_j(\tau)x_{j,t-h}$, for some $\tau \in [0,1]$. In order to obtain the result of the optimal forecast as an affine function of the conditional expectation, $\widetilde{f}_{i,t}^{h} = k_i^h + \beta_i^h\mathbb{E}_{t-h}(y_t)$, we need additional assumptions. The simplest way is to assume no scale effects, as follows:

In the absence of scale effects, only the intercept function $\alpha_0(\tau)$ varies across the quantile levels τ, that is $\alpha_j(\tau) = \overline{\alpha_j}$ for all $j = 1, ..., k$ and all $\tau \in [0,1]$. This way, the conditional quantile of y_t becomes $F_{t,t-h}^{-1}(\tau) = \alpha_0(\tau) + \sum_{j=1}^{k}\overline{\alpha_j}x_{j,t-h}$. Now, integrating both sides of the previous equation, and using the fact that integrating a conditional quantile function over the entire domain $\tau \in [0,1]$ yields the conditional mean (Koenker, 2005, p.302), gives:

$$\begin{aligned}\mathbb{E}_{t-h}(y_t) &= \int_0^1 F^{-1}_{t,t-h}(\tau)d\tau = \int_0^1 \left(\alpha_0(\tau) + \sum_{j=1}^{k}\overline{\alpha_j}x_{j,t-h}\right)d\tau \\ &\therefore \sum_{j=1}^{k}\overline{\alpha_j}x_{j,t-h} = \mathbb{E}_{t-h}(y_t) - \overline{\alpha_0},\end{aligned}$$

where $\overline{\alpha_0} = \int_0^1 \alpha_0(\tau)d\tau$. On the other hand, from the first part of the proof of Proposition 6, it follows that the optimal forecast $\widetilde{f}^h_{i,t}$ is a linear conditional quantile of y_t, evaluated at the optimal level τ^*, that is $\widetilde{f}^h_{i,t} = \alpha_0(\tau^*) + \sum_{j=1}^{k}\alpha_j(\tau^*)x_{j,t-h} = \alpha_0(\tau^*) + \sum_{j=1}^{k}\overline{\alpha_j}x_{j,t-h}$, where the last equality comes from the assumption of no scale effects. Now, we use the fact that $\sum_{j=1}^{k}\overline{\alpha_j}x_{j,t-h} = \mathbb{E}_{t-h}(y_t) - \overline{\alpha_0}$ to get the expression for the optimal forecast: $\widetilde{f}^h_{i,t} = \alpha_0(\tau^*) + \mathbb{E}_{t-h}(y_t) - \overline{\alpha_0} = (\alpha_0(\tau^*) - \overline{\alpha_0}) + \mathbb{E}_{t-h}(y_t) = k^h_i + \mathbb{E}_{t-h}(y_t)$. Therefore, note that $\widetilde{f}^h_{i,t} = k^h_i + \beta^h_i\mathbb{E}_{t-h}(y_t)$, where $\beta^h_i = 1$ (due to the absence of scale effects).

One way to relax the previous case (no scale effects) is to assume that the ratios $\alpha_j(\tau)/\alpha_1(\tau)$ are constant for all $\tau \in [0,1]$ and for all $j = 2, ..., k$, that is $\alpha_j(\tau)/\alpha_1(\tau) = w_j$. Note that $w_1 = 1$ and consider the conditional quantile of y_t as follows:

$$\begin{aligned}F^{-1}_{t,t-h}(\tau) &= \alpha_0(\tau) + \alpha_1(\tau)\left[x_{1,t-h} + \frac{\alpha_2(\tau)}{\alpha_1(\tau)}x_{2,t-h} + ... + \frac{\alpha_k(\tau)}{\alpha_1(\tau)}x_{k,t-h}\right] \\ &= \alpha_0(\tau) + \alpha_1(\tau)\left[w_1x_{1,t-h} + w_2x_{2,t-h} + ... + w_kx_{k,t-h}\right] \\ &= \alpha_0(\tau) + \alpha_1(\tau)\sum_{j=1}^{k}w_jx_{j,t-h}.\end{aligned}$$

Now, integrate both sides of the previous equation over the entire domain $\tau \in [0,1]$, in the following way:

$$\begin{aligned}\mathbb{E}_{t-h}(y_t) &= \int_0^1 F^{-1}_{t,t-h}(\tau)d\tau = \int_0^1 \alpha_0(\tau)d\tau + \int_0^1 \alpha_1(\tau)\sum_{j=1}^{k}w_jx_{j,t-h}d\tau \\ &\therefore \sum_{j=1}^{k}w_jx_{j,t-h} = \frac{\mathbb{E}_{t-h}(y_t) - \overline{\alpha_0}}{\overline{\alpha_1}}\end{aligned}$$

where $\overline{\alpha_0} = \int_0^1 \alpha_0(\tau)d\tau$ and $\overline{\alpha_1} = \int_0^1 \alpha_1(\tau)d\tau$. Now, evaluate the conditional quantile at the optimal level τ^* to get the optimal forecast, that is:

$$\begin{aligned}\widetilde{f}^h_{i,t} &= \alpha_0(\tau^*) + \alpha_1(\tau^*)\sum_{j=1}^{k}w_jx_{j,t-h} \\ &= \alpha_0(\tau^*) + \alpha_1(\tau^*)\left(\frac{\mathbb{E}_{t-h}(y_t) - \overline{\alpha_0}}{\overline{\alpha_1}}\right) = k^h_i + \beta^h_i\mathbb{E}_{t-h}(y_t),\end{aligned}$$

where $k^h_i = \left(\alpha_0(\tau^*) - \frac{\overline{\alpha_0}\alpha_1(\tau^*)}{\overline{\alpha_1}}\right)$ and $\beta^h_i = \frac{\alpha_1(\tau^*)}{\overline{\alpha_1}}$.

A.4 Data set of macroeconomic and financial variables

Table A.4 - Data set

Series	Category	Name	Source	Original	Transformation
1	Prices	IPCA (consumer price index)	IBGE	Index	% variation in 12 months
2	Prices	IPCA (consumer price index, market prices)	IBGE	Index	% variation in 12 months
3	Prices	IPCA (consumer price index, regulated and monitored prices)	IBGE	Index	% variation in 12 months
4	Prices	IPC-Fipe (consumer price index)	Fipe	Index	% variation in 12 months
5	Prices	IPC-Br (consumer price index)	FGV	Index	% variation in 12 months
6	Prices	IPA-DI (w holesale price index)	FGV	Index	% variation in 12 months
7	Prices	IGP-DI (general price index)	FGV	Index	% variation in 12 months
8	Prices	IGP-M (general price index)	FGV	Index	% variation in 12 months
9	Money	Monetary base	BCB	R$ thousand	% variation in 12 months
10	Money	Money supply (currency outside banks)	BCB	R$ thousand	% variation in 12 months
11	Money	Demand deposits	BCB	R$ thousand	% variation in 12 months
12	Money	Savings deposits	BCB	R$ thousand	% variation in 12 months
13	Money	M1	BCB	R$ thousand	% variation in 12 months
14	Money	M2	BCB	R$ thousand	% variation in 12 months
15	Money	M3	BCB	R$ thousand	% variation in 12 months
16	Money	M4	BCB	R$ thousand	% variation in 12 months
17	Interest rates	Nominal policy interest rate (Selic)	BCB	% p.a.	-
18	Interest rates	Nominal policy interest rate (long-term interest rate, TJLP)	BCB	% p.a.	-
19	Interest rates	Nominal market interest rate (Sw ap Pré-DI, 1 year)	Reuters	% p.a.	-
20	Interest rates	Real market interest rate (Sw ap Pré-DI, 1 year, deflator: Focus 12m inflation expect.)	Reuters, BCB	% p.a.	-
21	Banking sector	Credit spread (nonearmarked credit rate - Selic rate)	BCB	basis points	-
22	Banking sector	Non-Performing Loans (NPL) of total credit	BCB	%	-
23	Banking sector	Loan-to-Deposit ratio (LTD)	BCB	Units	-
24	Banking sector	Reserve requirements ratio (financial inst. reserve requirements / total deposits)	BCB	Units	-
25	Banking sector	Real grow th of nonearmarked credit operations outstanding	BCB	R$ million	% variation in 12 months
26	Capital markets	Ibovespa (Brazil)	Reuters	Index	% variation in 12 months
27	Capital markets	Initial Public Offers (IPOs) accumulated in 12 months (Brazil)	BCB	R$ million	% variation in 12 months
28	Capital markets	Net equity of stock funds (Brazil)	BCB	R$ million	% variation in 12 months
29	Capital markets	Net equity of financial investment funds (Brazil)	BCB	R$ million	% variation in 12 months
30	Capital markets	MSCI emerging countries (EM, US$)	Reuters	Index	% variation in 12 months
31	Capital markets	MSCI developed countries (World, US$)	Reuters	Index	% variation in 12 months
32	Risk premium	Embi+Br (Emerging Markets Bond Index Plus Brazil, spread)	Reuters	basis points	-
33	Risk premium	Embi+composite (average spread of 16 emerging countries)	Reuters	basis points	-
34	Risk premium	CDS (Credit Default Sw ap, Brazil 5 years)	Reuters	basis points	-
35	Exchange rates	FX-rate (nominal exchange rate, R$/US$)	BCB	Units	% variation in 12 months
36	Exchange rates	REER (Real effective exchange rate, IPA-13 currencies)	Funcex	Index	% variation in 12 months
37	Global Economy	U.S. dollar index (DXY, geometric average of 6 currencies in respect to US$)	Reuters	Index	% variation in 12 months
38	Global Economy	U.S. Treasury 2 years (Treasury nominal interest rates)	Reuters	% p.a.	-
39	Global Economy	U.S. Treasury 10 years (Treasury nominal interest rates)	Reuters	% p.a.	-
40	Global Economy	U.S. Treasury 5 years TIPS (Treasury Inflation-Protected Securities)	Reuters	% p.a.	-
41	Global Economy	CRB all commodities index	Reuters	Index	% variation in 12 months
42	Global Economy	Oil price (WTI, Oklahoma-USA)	Reuters	US$/barrel	% variation in 12 months
43	Global Economy	VIX CBOE volatility index (30-day expected volatility of the U.S. stock market, S&P500)	Reuters	Index	-
44	Exterior	Import price index	Funcex	Index	% variation in 12 months
45	Exterior	Import quantum index	Funcex	Index	% variation in 12 months
46	Exterior	Export price index	Funcex	Index	% variation in 12 months
47	Exterior	Export quantum index	Funcex	Index	% variation in 12 months
48	Exterior	Imports (FOB, total)	MDIC/Secex	US$	% variation in 12 months
49	Exterior	Exports (FOB, total)	MDIC/Secex	US$	% variation in 12 months
50	Exterior	Exports (FOB, primary goods)	MDIC/Secex	US$	% variation in 12 months
51	Exterior	International reserves (total)	BCB	US$ million	% variation in 12 months
52	Exterior	Current account (monthly, net)	BCB	US$ million	% variation in 12 months
53	Exterior	Current account (accumulated in 12 months, in relation to GDP)	BCB	%	-
54	Exterior	FDI (Foreign Direct Investment, accumulated in 12 months)	BCB	US$ million	% variation in 12 months
55	Exterior	FPI (Foreign Portfolio Investment, accumulated in 12 months)	BCB	US$ million	% variation in 12 months

Table A.4 - Data set (cont.)

Series	Category	Name	Source	Original	Transformation
56	Economic activity	IBC-BR (central bank economic activity index)	BCB	Index	% variation in 12 months
57	Economic activity	GDP (accumulated in the last 12 months, current prices)	BCB	R$ million	% variation in 12 months
58	Labor	Unemployment rate (open)	IBGE	%	-
59	Labor	Registered employees index (wholesale and retail trade)	MTE	Index	% variation in 12 months
60	Labor	Registered employees index (construction sector)	MTE	Index	% variation in 12 months
61	Labor	Hours worked in production (São Paulo)	Fiesp	Index	% variation in 12 months
62	Labor	Real overall wages (industry, São Paulo)	Fiesp	Index	% variation in 12 months
63	Industry	Industrial production (total)	IBGE	Index	% variation in 12 months
64	Industry	Industrial production (mineral extraction)	IBGE	Index	% variation in 12 months
65	Industry	Industrial production (manufacturing industry)	IBGE	Index	% variation in 12 months
66	Industry	Industrial production (capital goods)	IBGE	Index	% variation in 12 months
67	Industry	Industrial production (intermediate goods)	IBGE	Index	% variation in 12 months
68	Industry	Industrial production (consumer goods)	IBGE	Index	% variation in 12 months
69	Industry	Industrial production (durable goods)	IBGE	Index	% variation in 12 months
70	Industry	Industrial production (semidurable and nondurable goods)	IBGE	Index	% variation in 12 months
71	Industry	Installed capacity utilization (São Paulo)	Fiesp	%	seasonal adjustment
72	Industry	Capacity utilization (manufacturing industry, FGV)	FGV	%	seasonal adjustment
73	Industry	Vehicles production (total)	Anfavea	Units	% variation in 12 months
74	Industry	Passenger cars and light commercial vehicles production	Anfavea	Units	% variation in 12 months
75	Industry	Truck production	Anfavea	Units	% variation in 12 months
76	Industry	Bus production	Anfavea	Units	% variation in 12 months
77	Industry	Production of agricultural machinery (total)	Anfavea	Units	% variation in 12 months
78	Sales	Sales volume index in the retail sector (total)	IBGE	Index	% variation in 12 months
79	Sales	Sales volume index in the retail sector (fuel and lubricants)	IBGE	Index	% variation in 12 months
80	Sales	Sales volume index in the retail sector (hypermarkets, superm., food, bever. and tobacco)	IBGE	Index	% variation in 12 months
81	Sales	Sales volume index in the retail sector (textiles, clothing and footwear)	IBGE	Index	% variation in 12 months
82	Sales	Sales volume index in the retail sector (furniture and white goods)	IBGE	Index	% variation in 12 months
83	Sales	Sales volume index in the retail sector (vehicles and motorcycles, spare parts)	IBGE	Index	% variation in 12 months
84	Sales	Sales volume index in the retail sector (hypermarkets and supermarkets)	IBGE	Index	% variation in 12 months
85	Sales	Vehicle sales (total)	Anfavea	Units	% variation in 12 months
86	Sales	Domestic vehicle sales	Anfavea	Units	% variation in 12 months
87	Energy	Electric energy consumption (commercial)	Eletrobras	GWh	% variation in 12 months
88	Energy	Electric energy consumption (residential)	Eletrobras	GWh	% variation in 12 months
89	Energy	Electric energy consumption (industrial)	Eletrobras	GWh	% variation in 12 months
90	Energy	Electric energy consumption (other)	Eletrobras	GWh	% variation in 12 months
91	Energy	Electric energy consumption (total)	Eletrobras	GWh	% variation in 12 months
92	Public sector	Net public debt (% GDP, total, federal government and central bank)	BCB	%	-
93	Public sector	Net public debt (% GDP, internal, federal government and central bank)	BCB	%	-
94	Public sector	Net public debt (% GDP, external, federal government and central bank)	BCB	%	-
95	Public sector	Primary result of consolidated public sector (flow accumulated in 12 months, % GDP)	BCB	%	-