A Panel Data Approach to Economic Forecasting: The Bias-Corrected Average Forecast*

João Victor Issler
Luiz Renato Lima†
Graduate School of Economics – EPGE
Getulio Vargas Foundation
email: jissler@fgv.br and luizr@fgv.br

First Draft: December, 2006.

Abstract

In this paper, we propose a novel approach to econometric forecasting of stationary and ergodic time series within a panel-data framework. Our key element is to employ the bias-corrected average fore-

*We are especially grateful for the comments and suggestions given by Marcelo Fernandes, Antonio Galvão, Wagner Gaglianone, Raffaella Giacomini, Clive Granger, Marcelo Medeiros, Marcelo Moreira, Zhijie Xiao, and Hal White. We also benefited from comments given by the participants of the conference "Econometrics in Rio." We thank Wagner Gaglianone and Claudia Rodrigues for research assistance and gratefully acknowledge support given by CNPq-Brazil, CAPES, and Pronex. João Victor Issler thanks the hospitality of the Rady School of Management, and the Department of Economics of UCSD, where parts of this paper were written. Both authors thank the hospitality of University of Illinois, where the final revision was writen. The usual disclaimer applies.

†Corresponding Author.

cast. Using panel-data sequential asymptotics we show that it is potentially superior to other techniques in several contexts. In particular, it delivers a zero-limiting mean-squared error if the number of forecasts and the number of post-sample time periods is sufficiently large. We also develop a zero-mean test for the average bias. Monte-Carlo simulations are conducted to evaluate the performance of this new technique in finite samples. An empirical exercise, based upon data from well known surveys is also presented. Overall, these results show promise for the bias-corrected average forecast.

Keywords: Common Features, Panel-Data Econometrics.
J.E.L. Codes: C32, C33, E21, E44, G12.

1 Introduction

Bates and Granger(1969) made the econometric profession aware of the benefits of forecast combination when a limited number of forecasts is considered. The widespread use of different combination techniques has lead to an interesting puzzle from the econometrics point of view – the well known *forecast combination puzzle*: if we consider a fixed number of forecasts ($N < \infty$), combining them using equal weights ($1/N$) fare better than using "optimal weights" constructed to outperform any other forecast combination.

Regardless of how one combine forecasts, if the series being forecast is stationary and ergodic, and there is enough diversification among forecasts, we should expect that a weak law-of-large-numbers (WLLN) applies to well-behaved forecast combinations. Indeed, Timmermann(2006) uses financial-economic arguments based upon risk diversification to defend the idea of pooling of forecasts. This motivates labeling it "a financial approach to economic forecasts," since it is based on a principle so keen on finance; see, e.g., Ross (1976), Chamberlain and Rothschild (1983), and Connor and Korajzcyk (1986, 1993). Of course, to obtain this WLLN result, the number of forecasts has to diverge ($N \to \infty$), which entails the use of asymptotic panel-data techniques. In our view, one of the reasons why pooling forecasts has not yet been given a full asymptotic treatment is that forecasting is frequently thought to be a time-series experiment, not a panel-data experiment. As far as we know, despite its obvious benefits, there has been no work where the pooling forecasts was considered in a panel-data context, with the number of forecasts (N) and time-series observations (T) diverging without bounds.

In this paper, we propose a novel approach to econometric forecasting of stationary and ergodic series within a panel-data framework. First, we decompose individual forecasts into three components: the series being forecast, a time-invariant forecast bias, and a zero-mean forecast error. We show that the series being forecast is a *common feature* of all individual forecasts; see Engle and Kozicki(1993). Second, when $N, T \to \infty$, and we use standard tools from panel-data asymptotic theory, we show that the pooling of forecasts delivers optimal limiting forecasts in the sense that they have a zero mean-squared error. The key element of this result is the use of the *bias-corrected average forecast* – equal weights in combining forecasts coupled with a bias-correction term. The use of equal weights avoids estimating forecast weights, which contributes to reduce forecast variance, although potentially at the cost of an increase in bias. The use of a bias-correction term eliminates any possible detrimental effect arising from equal weighting. One important element of our technique is to use the forecast combination puzzle to our advantage, but now in an asymptotic context.

The use of the bias-corrected average forecast is a parsimonious choice in forecasting that delivers optimal forecasts in a mean-squared error sense – zero limiting mean-squared error. The only parameter we need to estimate is the mean bias, which requires the use the sequential asymptotic approach developed by Phillips and Moon (1999). Indeed, the only way we could increase parsimony in our framework is by doing without any bias correction. To test the usefulness of performing bias correction, we developed a zero-mean test for the average bias which draws upon the work of Conley (1999) on random fields.

Despite the lack of panel-data work on the pooling of forecasts, there has been panel-data research on forecasting focusing on pooling of information; see Stock and Watson (1999 and 2002a and b) and Forni et al. (2000, 2003). There, asymptotic theory was not used to pool forecasts, but information. The former is related to forecast combination and operates a reduction on the space of forecasts. The latter operates a reduction on a set of highly correlated regressors. In principle, forecasting can benefit from the use of both procedures. However, the payoff of pooling forecasts is greater than that of pooling information: while pooling information delivers optimal forecasts in the mean-squared error sense (Stock and Watson), it cannot drive the mean-squared forecast error to zero as the pooling of forecasts can.

One important element of our technique is the introduction of a bias-correction term. If a WLLN applies to a equal-weight forecast combination, we cannot guarantee a non-zero mean-squared error in forecasting, since the limit average bias of all forecasts may be non-zero. In this context, one interesting question that can be asked is the following: why are forecasts biased? From an economic standpoint, Laster, Bennett and Geoum (1999) show that professional forecasters behave strategically (i.e., they bias forecasts) if their payoffs depend mostly on publicity from the forecasts than from forecast-accuracy itself. Since one way to generate publicity is to deviate from a consensus (average) forecast, rewarding publicity may induce bias. From an econometric point of view, Patton and Timmermann (2006) consider an additional reason for the existence of bias in forecasts: what may look like forecast bias under a specific loss function may be just the consequence of the forecaster using a different loss function in producing the forecast[1]. Hoogstrate, Palm and Pfann (2000) show that pooling cross-sectional slopes can help in forecasting. One of the potential reasons why this procedure works in practice is that only cross-sectional slopes are pooled, not individual effects, showing that the latter may be working as a bias-correction device. A final

[1] Also, Clements and Hendry's (1999) work on intercept correction can be viewed as a study of bias.

reason for bias in forecasts is non-stationarity of the variable being forecast or of a subset of the conditioning variables. This is explored by Hendry and Clements (2002) and Clements and Hendry (2006).

Given that important forecast studies are motivated by bias in forecasting, it seems desirable to build a forecasting device that incorporates bias correction. We view the introduction of the bias-corrected average forecast as one of the original contributions of this paper. The way we estimate the bias-correction term relies on the use of a forecast-specific component to capture the bias in individual forecasts. Of course, this can only be fully studied asymptotically within a panel-data framework, which reinforces our initial choice of approach.

The ideas in this paper are related to research done in two 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), and the econometrics literature on forecasting related to common factors, forecast combination, bias correction, and structural breaks, perhaps best represented by the work of Bates and Granger (1969), Granger and Ramanathan(1984), Forni et al. (2000, 2003), Hendry and Clements (2002), Stock and Watson (2002a and b), Elliott and Timmermann (2003, 2004, 2005), and, more recently, by the excellent surveys of Clements and Hendry (2006), Stock and Watson (2006), and Timmermann (2006) – all contained in Elliott, Granger and Timmermann (2006). From finance and econometrics, our approach is related to the work on factor analysis when the number of assets is large, to recent work on panel-data asymptotics, and to panel-data methods focusing on financial applications, perhaps best exemplified by the work of Ross (1976), Chamberlain and Rothschild (1983), Connor and Korajzcyk (1986, 1993), Phillips and Moon (1999), Bai and Ng (2002, 2004), Bai (2005), and Pesaran (2005), and Araujo, Issler and Fernandes (2006).

The rest of the paper is divided as follows. Section 2 presents our main results and the assumptions needed to derive them. Proofs are presented in the Appendix. Section 3 presents the results of a Monte-Carlo experiment. Section 4 presents an empirical analysis using the methods proposed here, confronting the performance of our bias-corrected average forecast with that of other types of forecast combination. Section 5 concludes.

2 Econometric Setup and Main Results

Suppose that we are interested in forecasting a weakly stationary and ergodic univariate process $\{y_t\}$ using a large number of forecasts that will be combined to yield an optimal forecast in the mean-squared error (MSE) sense. These forecasts could be the result of using several econometric models that need to be estimated prior to forecasting, or the result of using no formal econometric model at all, e.g., just the result of an opinion poll on the variable in question using a large number of individual responses. We can also imagine that some (or all) of these poll responses are generated using econometric models, but then the econometrician that observes these forecasts has no knowledge of them.

We consider 3 consecutive distinct time periods, where time is indexed by $t = 1, 2, \ldots, T_1, \ldots, T_2, \ldots, T$. The period from $t = 1, 2, \ldots, T_1$ is labeled the "estimation sample," where models are usually fitted to forecast y_t, if that is the case. The period from $t = T_1 + 1, \ldots, T_2$ is labeled the post-model-estimation or "training sample", where realizations of y_t are usually confronted with forecasts produced in the estimation sample, if that is the case. The final period is $t = T_2 + 1, \ldots, T$, where genuine out-of-sample forecasting is entertained, benefiting from the results obtained during the training sample. In what follows, we let $T_2 \to \infty$. In order to guarantee that the number of observations in the training sample will go to infinity at rate T, we let T_1 be $O(1)$. Hence, asymptotic results will not hold for the estimation sample.

Regardless of whether forecasts are the result of a poll or of the estimation of an econometric model, we label forecasts of y_t, computed using conditioning sets lagged h periods, by $f_{i,t}^h$, $i = 1, 2, \ldots, N$. Therefore, $f_{i,t}^h$ are h-step-ahead forecasts and N is either the number of models estimated to forecast y_t or the number of respondents of an opinion poll regarding y_t.

West (1996) considers only two consecutive periods. The first has R data points are used to estimate models (R stands for regression), and the subsequent P data points are used for prediction. As opposed to West, in the case we have to estimate econometric models, we need 3 consecutive periods because we have two estimation periods. In the first we estimate models and in second these models are used to estimate a bias-correction term. Therefore, only in the third period genuine out-of-sample forecasting is entertained. In the case of surveys, since we do not have to estimate models, our setup is equivalent to that of West. Indeed, in his setup, $R, P \to \infty$ as $T \to \infty$, and $\lim\limits_{T\to\infty} R/P = \pi \in [0, \infty]$. Here, $R = T_2 - T_1$ and $P = T - T_2$, apart from irrelevant constants. Therefore,

$$\begin{aligned}
\pi &= \lim_{T\to\infty} R/P = \lim_{T\to\infty} \frac{T_2 - T_1}{T - T_2} = \lim_{T\to\infty} \frac{T_2/T - T_1/T}{1 - T_2/T} \\
&= \frac{\kappa}{1-\kappa}, \text{ where } 0 < \kappa < 1.
\end{aligned}$$

Hence, in our setup, $\pi \in (0, \infty)$ instead of as in West[2]. In what follows, to make our notation similar to that of West, we denote by R the number of observations in the training sample and by P the number of observations in the out-of-sample period.

In our setup, we also let N go to infinity, which raises the question of whether this is plausible in our context. On the one hand, if forecasts are the result of estimating econometric models, they will differ across i if they are either based upon different conditioning sets or upon different functional forms of the conditioning set (or both). Since there is an infinite number

[2] To inlcude the supports of $\pi \in [0, \infty]$ we must, asymptotically, give up having either a training sample or a genuine out-of-sample period.

of functional forms that could be entertained for forecasting, this gives an infinite number of possible forecasts. On the other hand, if forecasts are the result of a survey, although the number of responses is bounded from above, for all practical purposes, if a large enough number of responses is obtained, then the behavior of forecast combinations will be very close to the limiting behavior when $N \to \infty$.

Recall that, if we are interested in forecasting y_t, stationary and ergodic, using information up to h periods prior to t, then, under a MSE loss function, the optimal forecast is the conditional expectation using information available up to $t-h$: $\mathbb{E}_{t-h}(y_t)$. Using this well-known optimality result, Hendry and Clements (2002) argue that, the fact that the simple forecast average $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$ consistently outperforms individual forecasts $f_{i,t}^h$, shows the profession's inability to approximate $\mathbb{E}_{t-h}(y_t)$ reasonably well. With this motivation, our setup writes the $f_{i,t}^h$'s as approximations to the optimal forecast as follows:

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

where k_i is the individual model time-invariant bias and $\varepsilon_{i,t}$ is the indivual model error term in approximating $\mathbb{E}_{t-h}(y_t)$. Here, the optimal forecast is a common feature of all individual forecasts[3]. We can always decompose the series y_t into $\mathbb{E}_{t-h}(y_t)$ and an unforecastable componenet ζ_t, such that $\mathbb{E}_{t-h}(\zeta_t) = 0$ in:

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

Combining (1) and (2) yields,

$$\begin{aligned} f_{i,t}^h &= y_t - \zeta_t + k_i + \varepsilon_{i,t}, \text{ or,} \\ f_{i,t}^h &= y_t + k_i + \eta_t + \varepsilon_{i,t}, \text{ where, } \eta_t = -\zeta_t. \end{aligned} \tag{3}$$

[3]If an individual forecast is the conditional expectation $\mathbb{E}_{t-h}(y_t)$, then $k_i = \varepsilon_{i,t} = 0$. Notice that this implies that its MSE is smaller than that of $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$, something that is rarely seen in practice when a large number of forecasts are considered.

Equation (3) is indeed the well known two-way decomposition, or error-component decomposition, of the forecast error $f_{i,t}^h - y_t$:

$$\begin{aligned} f_{i,t}^h &= y_t + \mu_{i,t} \qquad i = 1, 2, \ldots, N, \qquad t > T_1, \\ \mu_{i,t} &= k_i + \eta_t + \varepsilon_{i,t}. \end{aligned} \tag{4}$$

It has been largerly used in econometrics dating back to Wallace and Hussein (1969), Amemiya (1971), Fuller and Battese (1974); see also Baltagi (1984) and Baltagi and Chang (1994) for a more recent setup. Davies and Lahiri (1995) used a three-way decomposition to investigate forecast rationality for the Survey of Professional Forecasts, within a panel-data context. As far as we know, neither a two- nor a three-way decomposition were used in studying the optimality of forecast combinations using panel-data asymptotics – the focus of our paper.

By construction, our framework in (4) specifies explict sources of forecast errors that are found in both y_t and $f_{i,t}^h$; see also the discussion in Davies and Lahiri. The term k_i is the time-invariant forecast bias of model i or of respondent i. It captures the long-run effect of forecast-bias of model i, or, in the case of surveys, the time invariant bias introduced by respondent i. Its source is $f_{i,t}^h$. The term η_t arises because forecasters do not have future information on y between $t-h+1$ and t. Hence, the source of η_t is y_t, and it is an additive aggregate zero-mean shock affecting equally all forecasts[4]. The term $\varepsilon_{i,t}$ captures all the remaining errors affecting forecasts, such as those of idiosyncratic nature and others that affect some but not all the forecasts (a group effect). Its source is $f_{i,t}^h$.

From equation (4), we conclude that k_i, $\varepsilon_{i,t}$ and η_t depend on the fixed horizon h. Here, however, to simplify notation, we do not make explicit this

[4] Because it is a component of y_t, and the forecast error is defined as $f_{i,t}^h - y_t$, the forecast error arising from lack of future information should have a negative sign in (4); see (3). To eliminate this negative sign, we defined η_t as the negative of this future-information component.

dependence on h. In our context, it makes sense to treat h as fixed and not as an additional dimension to i and t. In doing that, we follow West (1996) and the subsequent literature. As argued by Vahid and Issler (2002), forecasts are usually constructed for a few short horizons, since, as the horizon increases, the MSE in forecasting gets hopelessly large. Here, h will not vary as much as i and t, especially because $N, T \rightarrow \infty$[5].

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

We now list our set of assumptions.

Assumption 1 We assume that k_i, $\varepsilon_{i,t}$ and η_t are independent of each other for all i and t.

Independence is an algebraically convenient assumption used throughout the literature on two-way decompositions; see Wallace and Hussein (1969) and Fuller and Battese (1974) for example. At the cost of unnecessary complexity, it could be relaxed to use orthogonal components, e.g., Baltagi (1980) and the subsequent literature, something we avoid here.

Assumption 2 k_i is an identically distributed random variable in the cross-sectional dimension, but not necessarily independent, i.e.,

$$k_i \sim \text{i.d.}(B, \sigma_k^2), \tag{5}$$

[5] Davies and Lahiri considered a three-way decomposition with h as an added dimension. The foucs of their paper is forecast rationality. In their approach, η_t and $\varepsilon_{i,t}$ depend on h but k_i does not, the latter being critical to identify k_i within their framework. Since, in general, this restriction does not have to hold, our two-way decomposition is not nested into their three-way decompostion. Indeed, in our approach, k_i varies with h and it is still identified. We leave treatment of a varying horizon, within our framework, for future research.

where B and σ_k^2 are respectively the mean and variance of k_i. In the time-series dimension, k_i has no variation, therefore, it is a fixed parameter.

The idea of dependence is consistent with the fact that forecasters learn from each other by meeting, discussing, debating, etc. Through their ongoing interactions, they maintain a current collective understanding of where their target variable is most likely heading to and its upside and downside risks. Given the assumption of identical distribution for k_i, B represents the market (or collective) bias. Since we focus on combining forecasts, a pure idiosyncratic bias does not matter but a collective bias does. In principle, we could allow for heterogeneity in the distribution of k_i – means and variances to differ across i. However, that will be a problem in testing the hypothesis that forecast combinations are biased.

It is desirable to discuss the nature of the term k_i, which is related to the question of why we cannot focus solely on unbiased forecasts, for which $k_i = 0$. The role of k_i is to capture the long-run effect, in the time dimension, of the bias of econometric models of y_t, or of the bias of respondent i. A relevant question to ask is – why would forecasters introduce bias under MSE loss? Laster, Bennett and Geoum (1999), Patton and Timmermann (2006), and Batchelor (2007) list different arguments consistent with forecasters having a loss function different from MSE. The argument applies for surveys and for models as well, since a forecaster can use a model that is unbiased and add a bias term to it. In the examples that follow, all forecasters employ a combination of MSE loss and a secondary loss function. Bias is simply a consequence of this secondary loss function and of the intensity in which the forecaster cares for it. The first example is that of a bank selling an investment fund. In this case, the bank's forecast of the fund return may be upward-biased simply because it may use this forecast as a marketing strategy to attract new clients for that fund. Although the bank is penalized by deviating from $\mathbb{E}_{t-h}(y_t)$, it also cares for selling the shares of its fund. The second example introduces bias when there is a market for *pessimism* or *optimism* in forecasting. Forecasters may want to be labelled as optimists or pessimists in a "branding" strategy to be experts on "worst-" or on "best-case scenarios," respectively. Batchelor lists governments as examples of experts on the latter.

Assumption 3 The aggregate shock η_t is a stationary and ergodic MA process of order at most $h-1$, with zero mean and variance $\sigma_\eta^2 < \infty$.

Since h is a bounded constant in our setup, η_t is the result of a cumulation of shocks to y_t that occurred between $t-h+1$ and t. Being an $MA\,(h-1)$ is a consequence of the wold representation for y_t and of (2). If y_t is already an $MA\,(\cdot)$ process, of order smaller than $h-1$, then, its order will be the same of that of η_t. In any case, it must be stressed that η_t is unpredictable, i.e., that $\mathbb{E}_{t-h}\,(\eta_t) = 0$. This a consequence of (2) and of the law of iterated expectations, simply showing that, from the perspective of the forecast horizon h, unless the forecaster has superior information, the aggregate shock η_t cannot be predicted.

Assumption 4: Let $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots \varepsilon_{N,t})'$ be a $N \times 1$ vector stacking the errors $\varepsilon_{i,t}$ associated with all possible forecasts. Then, the vector process $\{\boldsymbol{\varepsilon}_t\}$ is assumed to be covariance-stationary and ergodic for the first and second moments, uniformly on N. Further, defining as $\xi_{i,t} = \varepsilon_{i,t} - \mathbb{E}_{t-1}\,(\varepsilon_{i,t})$, the innovation of $\varepsilon_{i,t}$, we assume that

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| = 0. \tag{6}$$

Non-egordicity of ε_t would be a consequence of the forecasts $f_{i,t}^h$ beyond k_i. Of course, forecasts that imply a non-ergodic ε_t could be discarded. Because the forecasts are computed h-steps ahead, forecast errors $\varepsilon_{i,t}$ can be serially correlated. Assuming that $\varepsilon_{i,t}$ is weakly stationary is a way of controlling its time-series dependence. It does not rule out errors displaying conditional heteroskedasticity, since the latter can coexist with the assumption of weak stationarity; see Engle (1982) and Bollerslev (1986).

Equation (6) limits the degree of cross-sectional dependence of the errors $\varepsilon_{i,t}$. It allows cross-correlation of the form present in a specific group of forecasts, although it requires that this cross-correlation will not prevent a weak law-of-large-numbers from holding. Following the forecasting literature with large N and T, e.g., Stock and Watson (2002b), and the financial econometric literature, e.g., Chamberlain and Rothschild (1983), the condition $\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right) \right| = 0$ controls the degree of cross-sectional decay in forecast errors. It is noted by Bai (2005, p. 6), that Chamberlain and Rothschild's cross-sectional error decay requires:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right) \right| < \infty. \tag{7}$$

Notice that this is the same cross-sectional decay used in Stock and Watson. Of course, (7) implies (6), but the converse is not true. Hence, Assumption 2 has a less restrictive condition than those commonly employed in the literature of factor models.

We state now basic results related to the classic question of "to pool or not to pool forecasts," when only simple weights ($1/N$) are used; see, for example, Granger (1989) and Palm and Zellner (1992).

Proposition 1 *Under Assumptions 1-4, the mean-squared error in forecasting y_t, using the individual forecast $f_{i,t}^h$, is $\mathbb{E}\left(f_{i,t}^h - y_t\right)^2 = k_i^2 + \sigma_\eta^2 + \sigma_{\epsilon_i}^2$, where $\sigma_{\epsilon_i}^2$ is the variance of $\varepsilon_{i,t}$, $i = 1, 2, , ..., N$.*

Proof. Start with:

$$f_{i,t}^h - y_t = k_i + \eta_t + \varepsilon_{i,t}.$$

Then,

$$\begin{aligned} MSE_i &= \mathbb{E}\left(f_{i,t}^h - y_t\right)^2 = \mathbb{E}(k_i + \eta_t + \varepsilon_{i,t})^2 = \mathbb{E}\left(k_i^2\right) + \mathbb{E}\left(\eta_t^2\right) + \mathbb{E}\left(\varepsilon_{i,t}^2\right) \\ &= k_i^2 + \sigma_\eta^2 + \sigma_{\epsilon_i}^2 \end{aligned} \tag{8}$$

where $\sigma^2_{\epsilon_i}$ is the variance of $\varepsilon_{i,t}$. Assumption 1 is used in the last line of (8). We also use the fact that k_i is a constant in the time-series dimension in the last line of (8). ■

Proposition 2 *Under Assumptions 1-4, as $N \to \infty$, the mean-squared error in forecasting y_t, combining all possible individual forecasts $f^h_{i,t}$, is*

$$MSE_{average} = \mathbb{E}\left(\underset{N\to\infty}{plim}\frac{1}{N}\sum_{i=1}^{N} f^h_{i,t} - y_t\right)^2 = B^2 + \sigma^2_\eta.$$

Proof. Start with the cross-sectional average of (4):

$$\frac{1}{N}\sum_{i=1}^{N} f^h_{i,t} - y_t = \frac{1}{N}\sum_{i=1}^{N} k_i + \eta_t + \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}.$$

Computing the probability limit of the right-hand side above gives,

$$\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N} k_i + \eta_t + \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}\right) = \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N} k_i + \eta_t + \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}. \tag{9}$$

We will compute the probability limits in (9) separately. The first one is a strightforward application fo the law of large numbers:

$$\underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N} k_i = B.$$

The second will turn out to be zero. Our strategy is to show that, in the limit, the variance of $\frac{1}{N}\sum\limits_{i=1}^{N}\varepsilon_{i,t}$ is zero, a sufficient condition for a weak law-of-large-numbers (WLLN) to hold for $\{\varepsilon_{i,t}\}_{i=1}^{N}$.

Because $\varepsilon_{i,t}$ is weakly stationary and mean-zero, for every i, there exists a scalar Wold representation of the form:

$$\varepsilon_{i,t} = \sum_{j=0}^{\infty} b_{i,j}\xi_{i,t-j} \tag{10}$$

where, for all i, $b_{i,0} = 1$, $\sum_{j=0}^{\infty} b^2_{i,j} < \infty$, and $\xi_{i,t}$ is white noise.

In computing the variance of $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{\infty} b_{i,j}\xi_{i,t-j}$ we use the fact that there is no cross correlation between $\xi_{i,t}$ and $\xi_{i,t-k}$, $k = 1, 2, \ldots$. Therefore, we need only to consider the sum of the variances of terms of the form $\frac{1}{N}\sum_{i=1}^{N} b_{ik}\xi_{i.t-k}$. These variances are given by:

$$\mathrm{VAR}\left(\frac{1}{N}\sum_{i=1}^{N} b_{i,k}\xi_{i,t-k}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} b_{i,k}b_{j,k}\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right), \tag{11}$$

due to weak stationarity of ε_t. We now examine the limit of the generic term in (11) with detail:

$$\mathrm{VAR}\left(\frac{1}{N}\sum_{i=1}^{N} b_{i,k}\xi_{i,t-k}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} b_{i,k}b_{j,k}\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right) \leq$$

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \left|b_{i,k}b_{j,k}\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \left|b_{i,k}b_{j,k}\right|\left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| \leq \tag{12}$$

$$\left(\max_{i,j}\left|b_{i,k}b_{j,k}\right|\right)\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right|. \tag{13}$$

Hence:

$$\begin{aligned}\lim_{N\to\infty}\mathrm{VAR}\left(\frac{1}{N}\sum_{i=1}^{N} b_{i,k}\xi_{i,t-k}\right) &\leq \lim_{N\to\infty}\left(\max_{i,j}\left|b_{i,k}b_{j,k}\right|\right)\times \\ \lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| &= 0,\end{aligned}$$

since the sequence $\{b_{i,j}\}_{j=0}^{\infty}$ is square-summable, yielding $\lim_{N\to\infty}\left(\max_{i,j}\left|b_{i,k}b_{j,k}\right|\right) < \infty$, and Assumption 4 imposes $\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\mathbb{E}\left(\xi_{i,t}\xi_{j,t}\right)\right| = 0$.

Thus, all variances are zero in the limit, as well as their sum, which gives:

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} = 0.$$

Therefore,

$$\mathbb{E}\left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h}-y_t\right)^2 = \mathbb{E}\left(B+\eta_t\right)^2$$
$$= B^2+\sigma_\eta^2. \tag{14}$$

■

We can now compare the MSE of a generic individual forecast with that of an equally weighted ($1/N$) forecast combination by using the usual bias-variance standard decomposition of the mean squared error (MSE)

$$MSE = Bias^2 + VAR.$$

Proposition 1 shows that we can decompose individual MSE's, MSE_i, as:

$$\begin{aligned} MSE_i &= k_i^2+\sigma_\eta^2+\sigma_{\epsilon_i}^2 \\ &= Bias_i^2+VAR_i,\ i=1,2,...,N \end{aligned}$$

where $Bias_i^2 = k_i^2$ and $VAR_i = \sigma_\eta^2+\sigma_{\epsilon_i}^2$.

Proposition 2 shows that averaging forecasts reduces variance, but not necessarily MSE,

$$\begin{aligned} MSE_{average} &= B^2+\sigma_\eta^2 \qquad (15) \\ &= Bias_{average}^2+VAR_{average}, \end{aligned}$$

where $VAR_{average} = \sigma_\eta^2 < VAR_i = \sigma_\eta^2+\sigma_{\epsilon_i}^2$, but comparing $Bias_{average}^2 = B^2$ with $Bias_i^2 = k_i^2$ requires knowledge of B and k_i, which is also true for comparing $MSE_{average}$ with MSE_i. If the mean bias $B=0$, i.e., we are considering unbiased forecasts, on average, then $MSE_i = k_i^2+\sigma_\eta^2+\sigma_{\epsilon_i}^2$, while $MSE_{average} = \sigma_\eta^2$. Therefore, if the number of forecasts in the combination is large enough, combining forecasts with a zero collective bias will lead to a smaller MSE – as concluded in Granger (1989). However, if $B \neq 0$, we cannot conclude that the average forecast has MSE lower than that of individual forecasts, since B^2 may be larger or smaller than $k_i^2 + \sigma_{\epsilon_i}^2$.

This motivates studying bias-correction in forecasting, since one way to eliminate the term B^2 in (14) is to perform bias correction coupled with equal weights $(1/N)$ in the forecast combination. The next set of results investigates the properties of the bias-corrected average forecast.

Proposition 3 *If Assumptions 1-4 hold, then, the bias-corrected average forecast, given by* $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i$, *obeys* $\underset{N\to\infty}{plim}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i\right) = y_t + \eta_t$ *and has a mean-squared error as follows:*
$MSE_{BCAF} = \mathbb{E}\left[\underset{N\to\infty}{plim}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i\right) - y_t\right]^2 = \sigma_\eta^2$. *Therefore, it is an optimal forecasting device in the MSE sense.*

Proof. From the proof of Propostion 2, we have:

$$\begin{aligned}\underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - y_t - \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N} k_i &= \eta_t + \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t} \\ &= \eta_t,\end{aligned}$$

leading to:

$$\mathbb{E}\left[\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \frac{1}{N}\sum_{i=1}^{N} k_i\right) - y_t\right]^2 = \sigma_\eta^2.$$

■

Proposition 3 shows that the bias-corrected average forecast is an optimal forecast in the MSE sense. Bias correction eliminates the term B^2 from the MSE expression, while equal weights naturally eliminates the variance of idiosyncratic components and group effects. The only term left in the MSE is σ_η^2, related to unforecastable news to the target variable after the forecast combination was computed – something we could not eliminate unless we had superior (future) information. From a finance perspective, all risks associated with terms that could be diversified were eliminated by using the bias-corrected average forecast. We were left only with the undiversifiable risk expressed in σ_η^2. Therefore, the optimal result.

There are infinite ways of combining forecasts. So far, we have considered only equal weights $1/N$. In order to discuss the forecast-combination puzzle, we now consider other combination schemes, consistent with a weak law-of-large-numbers (WLLN) for forecast combinations, i.e., bounded weights that add up to unity, in the limit.

Corollary 4 *Consider the sequence of deterministic weights $\{\omega_i\}_{i=1}^N$, such that $|\omega_i| < \infty$ uniformly on N and $\lim_{N\to\infty} \sum_{i=1}^{N} \omega_i = 1$. Then, under Assumptions 1-4,*

$$\begin{aligned} \underset{N\to\infty}{plim} \left(\sum_{i=1}^{N} \omega_i f_{i,t}^h - \sum_{i=1}^{N} \omega_i k_i - y_t \right) &= \eta_t, \text{ and,} \\ \mathbb{E}\left[\underset{N\to\infty}{plim} \left(\sum_{i=1}^{N} \omega_i f_{i,t}^h - \sum_{i=1}^{N} \omega_i k_i \right) - y_t \right]^2 &= \sigma_\eta^2. \end{aligned}$$

and the same result of Proposition 3 follows when a generic $\{\omega_i\}_{i=1}^N$ is used instead of $1/N$.

This corollary to Proposition 3 shows that there is not a unique optimum in the MSE sense. Indeed, any other combination scheme consistent with a WLLNs will be optimal as well. Of course, "optimal" population weights, constructed from the variance-covariance structure of models with stationary data, will obey the structure in Corollary 4. Hence, "optimal" population weights canot perform better than $1/N$ under bias correction. Therefore, there is no forecast-combination puzzle in the context of populational weights.

Although the discussion using populational weights is useful, the puzzle is associated with weights ω_i estimated using data. Therefore, we now compare $\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h$ with a bias-corrected version of $\sum_{i=1}^{N} \omega_i f_{i,t}^h$ with estimated weights.

We follow the discussion in Hendry and Clements (2002) using N different forecasts instead of just 2. Weights ω_i can be estimated ($\widehat{\omega}_i$) by running the following regression, minimizing MSE subject to $\sum_{i=1}^{N} \omega_i = 1$:

$$y = \delta \mathbf{i} + \omega_1 f_1 + \omega_2 f_2 + ... + \omega_N f_N + \upsilon, \tag{16}$$

where y denotes the $R \times 1$ vector of observations of the target variable, $f_1,\ f_2, ..., f_N$ denotes, respectively, the $R \times 1$ vectors of observations of the N individual forecasts, and $\mathbf{i}$ is a vector of ones. Estimation is done over the time interval $T_1 + 1, \ldots, T_2$ (i.e., over the training sample). On the one hand, because regression (16) includes an intercept, the forecast $\widehat{f} = \widehat{\delta}\mathbf{i} + \widehat{\omega}_1 f_1 + \widehat{\omega}_2 f_2 + ... + \widehat{\omega}_N f_N$ is unbiased, but its variance grows with N, since we have to estimate N weights to construct it. Notice that δ plays the role of a bias-correction term.

There are two cases to be considered. The behavior of estimated weights in small samples and asymptotically, when $N, T \to \infty$. In both cases, feasible estimates requires $N < R$. In small samples, when N is close to R from below, the variance of $\widehat{f}$ may be big enough as to yield an inferior forecast (MSE) relative to $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$, although the latter is biased. Thus, the weighted forecast cannot avoid the "curse of dimensionality" that plagues several estimates across econometrics. In this context, the curse of dimensionality in $\widehat{f}$ is a possible explanation to the forecast-combination puzzle[6]. Asymptotically, feasibility requires:

$$0 < \lim_{N,T \to \infty} \frac{N}{R} = c < 1, \tag{17}$$

which implies that $N \to \infty$ at a smaller rate than T[7]. As long as this

[6] We thank an anonymous referee for casting the problem in these terms.

[7] We could cast this condition in terms of N, T_2 and T_1. Then,

$$0 < \lim_{N,T \to \infty} \frac{N}{R} = \lim_{N,T \to \infty} \frac{N}{T_2 - T_1} = \lim_{N,T \to \infty} \frac{N/T}{T_2/T - T_1/T} = \frac{\kappa_1}{\kappa} = c < 1,$$

which requires that $N \to \infty$ at a smaller rate than T_2.

condition is achieved, weights are estimated consistently in (16) and we are back to Corollary 4 – asymptotically, there is no forecast-combination puzzle.

The bias-variance tradeoff in MSE motivates the main question of our paper as follows: can we compute a forecast combination that will have asymptotically the same variance as in $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}$ and zero bias as in $\widehat{f}$? It turns out that the answer is yes – the bias-corrected average forecast (BCAF) in Proposition 4. Hence, we are able to improve upon the simple average forecast[8].

Despite the optimal behavior of the bias-corrected average forecast $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N} k_i$, it is imediately seen that it is unfeasible because the k_i's are unknown. Therefore, below, we propose replacing k_i by a consistent estimator. The underlying idea behind the consistent estimator of k_i is that, in the training sample, one observes the realizations of y_t and $f_{i,t}^{h}$, $i = 1...N$, for the R training-sample observations. Hence, one can form a panel of forecasts:

$$\left(f_{i,t}^{h} - y_t\right) = k_i + \eta_t + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad t = T_1 + 1, \cdots, T_2, \quad (18)$$

where it becomes obvious that k_i represents the fixed effect on this panel. It is natural to exploit this property of k_i in constructing a consistent estimator. This is exactly the approach taken here. In what follows, we propose a non-parametric estimator of k_i. It does not depend on any distributional assumption on $k_i \sim \text{i.d.}(B, \sigma_k^2)$ and it does not depend on any knowledge of the models used to compute the forecasts $f_{i,t}^{h}$. This feature of our approach widens its application to situations where the "underlying models are not known, as in a survey of forecasts," as discussed by Kang (1986).

Due to the nature of our problem – large number of forecasts – and the

[8] Only in an asymptotic panel-data framework can we formally state weak law-of-large-numbers for forecast combinations. We see this as a major advantage of our approach vis-à-vis the commonly employed time-series approach with fixed N – especially when $N = 2$ or $N = 3$.

nature of k_i in (18) – time-invariant bias term – we need to consider large N, large T asymptotic theory to devise a consistent estimator for k_i. Panels with such a character are different from large N, small T panels. In order to allow the two indices N and T to pass to infinity jointly, we could consider a monotonic increasing function of the type $T = T(N)$, known as diagonal-asymptotic method; see Quah (1994) and Levin and Lin (1993). One drawback of this approach is that the limit theory that is obtained depends on the specific relationship considered in $T = T(N)$. A joint-limit theory allows both indices (N and T) to pass to infinity simultaneously without imposing any specific functional-form restrictions. Despite that, it is substantially more difficult to derive and will usually apply only under stronger conditions, such as the existence of higher moments. Searching for a method that allows robust asymptotic results without imposing too many restrictions (on functional relations and the existence of higher moments), we consider the sequential asymptotic approach developed by Phillips and Moon (1999). There, one first fixes N and then allows T to pass to infinity using an intermediate limit. Phillips and Moon write sequential limits of this type as $(T, N \to \infty)_{\text{seq}}$.

By using the sequential-limit approach of Phillips and Moon, we now show how to estimate k_i, B, and η_t consistently.

Proposition 5 *If Assumptions 1-4 hold, the following are consistent estimators of k_i, B, η_t, and $\varepsilon_{i,t}$, respectively:*

$$\begin{aligned}
\widehat{k}_i &= \frac{1}{R}\textstyle\sum_{t=T_1+1}^{T_2} f_{i,t}^h - \frac{1}{R}\sum_{t=T_1+1}^{T_2} y_t, & \underset{T\to\infty}{plim}\left(\widehat{k}_i - k_i\right) = 0, \\
\widehat{B} &= \frac{1}{N}\textstyle\sum_{i=1}^{N} \widehat{k}_i, \text{ and,} & \underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\widehat{B} - B\right) = 0, \\
\widehat{\eta}_t &= \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h - \widehat{B} - y_t, & \underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\widehat{\eta}_t - \eta_t\right) = 0, \\
\widehat{\varepsilon}_{i,t} &= f_{i,t}^h - y_t - \widehat{k}_i - \widehat{\eta}_t, & \underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\widehat{\varepsilon}_{i,t} - \varepsilon_{i,t}\right) = 0.
\end{aligned}$$

Proof. Although y_t, η_t and $\varepsilon_{i,t}$ are ergodic for the mean, $f_{i,t}^h$ is non ergodic because of k_i. Recall that, $T_2, R \to \infty$, as $T \to \infty$. As $T \to \infty$,

$$
\begin{aligned}
\frac{1}{R}\sum_{t=T_1+1}^{T_2} f_{i,t}^h &= \frac{1}{R}\sum_{t=T_1+1}^{T_2} y_t + \frac{1}{R}\sum_{t=T_1+1}^{T_2} \varepsilon_{i,t} + \frac{1}{R}\sum_{t=T_1+1}^{T_2} \eta_t + k_i \\
&\xrightarrow{p} \mathbb{E}(y_t) + k_i + \mathbb{E}(\varepsilon_{i,t}) + \mathbb{E}(\eta_t) \\
&= \mathbb{E}(y_t) + k_i
\end{aligned}
$$

Given that we observe $f_{i,t}^h$ and y_t, we propose the following consistent estimator for k_i, as $T \to \infty$:

$$
\begin{aligned}
\widehat{k}_i &= \frac{1}{R}\sum_{t=T_1+1}^{T_2} f_{i,t}^h - \frac{1}{R}\sum_{t=T_1+1}^{T_2} y_t, \quad i = 1, ..., N \\
&= \frac{1}{R}\sum_{t=T_1+1}^{T_2} (y_t + k_i + \eta_t + \varepsilon_{i,t}) - \frac{1}{R}\sum_{t=T_1+1}^{T_2} y_t \\
&= k_i + \frac{1}{R}\sum_{t=T_1+1}^{T_2} \varepsilon_{i,t} + \frac{1}{R}\sum_{t=T_1+1}^{T_2} \eta_t \\
&\text{or,} \\
\widehat{k}_i - k_i &= \frac{1}{R}\sum_{t=T_1+1}^{T_2} \varepsilon_{i,t} + \frac{1}{R}\sum_{t=T_1+1}^{T_2} \eta_t.
\end{aligned}
$$

Using this last result, we can now propose a consistent estimator for B:

$$
\widehat{B} = \frac{1}{N}\sum_{i=1}^{N} \widehat{k}_i = \frac{1}{N}\sum_{i=1}^{N} \left[\frac{1}{R}\sum_{t=T_1+1}^{T_2} f_{i,t}^h - \frac{1}{R}\sum_{t=T_1+1}^{T_2} y_t\right].
$$

First let $T \to \infty$,

$$
\begin{aligned}
&\widehat{k}_i \xrightarrow{p} k_i, \text{ and,} \\
&\frac{1}{N}\sum_{i=1}^{N} \widehat{k}_i \xrightarrow{p} \frac{1}{N}\sum_{i=1}^{N} k_i.
\end{aligned}
$$

Now, as $N \to \infty$, after $T \to \infty$,

$$
\frac{1}{N}\sum_{i=1}^{N} k_i \xrightarrow{p} B,
$$

Hence, as $(T, N \to \infty)_{\text{seq}}$,

$$
\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}} \left(\widehat{B} - B\right) = 0.
$$

We can now propose a consistent estimator for η_t:

$$\widehat{\eta}_t = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \widehat{B} - y_t = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N} \widehat{k}_i - y_t.$$

We let $T \to \infty$ to obtain:

$$\begin{aligned} \operatorname*{plim}_{T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i - y_t\right) &= \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \frac{1}{N}\sum_{i=1}^{N} k_i - y_t \\ &= \eta_t + \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}. \end{aligned}$$

Letting now $N \to \infty$ we obtain $\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} = 0$ and:

$$\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}} \left(\widehat{\eta}_t - \eta_t\right) = 0.$$

Finally,

$$\begin{aligned} \widehat{\varepsilon}_{i,t} &= f_{i,t}^{h} - y_t - \widehat{k}_i - \widehat{\eta}_t, \text{ and } f_{i,t}^{h} - y_t = k_i + \eta_t + \varepsilon_{i,t}. \\ \text{Hence} &: \\ \widehat{\varepsilon}_{i,t} - \varepsilon_{i,t} &= \left(k_i - \widehat{k}_i\right) + \left(\eta_t - \widehat{\eta}_t\right). \end{aligned}$$

Using the previous results that $\operatorname*{plim}_{T\to\infty}\left(\widehat{k}_i - k_i\right) = 0$ and $\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}}\left(\widehat{\eta}_t - \eta_t\right) = 0$, we obtain:

$$\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}} \left(\widehat{\varepsilon}_{i,t} - \varepsilon_{i,t}\right) = 0.$$

■

The result above shows how to construct feasible estimators in a sequential asymptotic framework, leading to the feasible bias-corrected average forecast. We now state our most important result.

Proposition 6 *If Assumptions 1-4 hold, the feasible bias-corrected average forecast* $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \widehat{B}$ *obeys* $\operatorname*{plim}_{(T,N\to\infty)_{\text{seq}}}\left(\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h} - \widehat{B}\right) = y_t + \eta_t = \mathbb{E}_{t-h}\left(y_t\right)$

and has a mean-squared error as follows:
$\mathbb{E}\left[\underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h}-\widehat{B}\right)-y_t\right]^2=\sigma_\eta^2$. *Therefore it is an optimal forecasting device.*

Proof. We let $T\to\infty$ to obtain:

$$\begin{aligned}\underset{T\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h}-\widehat{B}\right) &= \underset{T\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h}-\frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i\right)\\ &\underset{T\to\infty}{}\\ = \quad \underset{T\to\infty}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h}-\frac{1}{N}\sum_{i=1}^{N}k_i\right) &= y_t+\eta_t+\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}.\end{aligned}$$

Letting now $N\to\infty$ we obtain $\underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}=0$ and:

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

which is the optimal forecast, from (3). The MSE of the feasible bias-corrected average forecast is:

$$\mathbb{E}\left[\underset{(T,N\to\infty)_{\text{seq}}}{\text{plim}}\left(\frac{1}{N}\sum_{i=1}^{N}f_{i,t}^{h}-\widehat{B}\right)-y_t\right]^2=\sigma_\eta^2.$$

showing that we are back to the result in Proposition 3. ■

Here, combining forecasts using equal weights $1/N$ and bias correction is also optimal, and we can approximate $\mathbb{E}_{t-h}(y_t)$ well enough. As before, any other forecast combination as in Corollary 4 will also be optimal. Again, there is no forecast combination puzzle here.

The advantage of equal weights $1/N$ is not having to estimate weights. To get optimal forecasts, in the MSE sense, one has to combine all forecasts using simple averaging, appropriately centering it by using a bias-correction term. It is important to stress that, even though $N\to\infty$, the number of estimated parameters is kept at unity: $\widehat{B}$. This is a very attractive feature of our approach compared to models that combine forecasts estimating "optimal" weights, where the number of estimated parameters increases at the same rate as N. Our answer to the curse of dimensionality is parsimony, implied by estimating only one parameter – $\widehat{B}$[9].

Parsimony can be viewed form a different angle. From (16), we could retrieve $\widehat{B}$ from an OLS regression of the form:

$$y = \delta \mathbf{i} + \omega_1 f_1 + \omega_2 f_2 + ... + \omega_N f_N + \upsilon,$$

where the weights ω_i are constrained to be $\omega_i = 1/N$ for all i. There is only one parameter to be estimated, δ, and $\widehat{\delta} = \widehat{B}$, where $\widehat{B}$ is now cast in terms of the previous literature.

The feasible bias-corrected average forecast can be made an even more parsimonious estimator of y_t when there is no need to estimate B. Of course, this raises the issue of whether $B = 0$, in which case the optimal forecast becomes $\frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$ – the simple forecast combination originally proposed by Bates and Granger (1969). We next propose the following test statistic for $H_0 : B = 0$.

Proposition 7 *Under the null hypothesis $H_0 : B = 0$, the test statistic:*

$$\widehat{t} = \frac{\widehat{B}}{\sqrt{\widehat{V}}} \quad \xrightarrow[(T,N\to\infty)_{\text{seq}}]{d} \mathcal{N}(0,1),$$

[9]From a different perspective, notice that, $\widehat{B} = \frac{1}{N}\sum_{i=1}^{N} \widehat{k}_i$, where each $\widehat{k}_i$ is estimated separately using R observations, whereas, $\{\widehat{\omega}_i\}_{i=1}^{N}$ is jointly estimated using these same R observations. Our estimator is also less restrictive form an asymptotic point-of-view. For the weighted forecast combination, recall that the feasibility condition required that,

$$0 < \lim_{N,T\to\infty} \frac{N}{R} = c < 1.$$

Here, $0 < c < \infty$.

where $\widehat{V}$ *is a consistent estimator of the asymptotic variance of* $\overline{B} = \frac{1}{N}\sum_{i=1}^{N} k_i$.

Proof. Under $H_0 : B = 0$, we have shown in Proposition 3 that $\widehat{B}$ is a $(T, N \to \infty)_{\text{seq}}$ consistent estimator for B. To compute the consistent estimator of the asymptotic variance of $\overline{B}$ we follow Conley(1999), who matches spatial dependence to a metric of economic distance. Denote by $\text{MSE}_i(\cdot)$ and $\text{MSE}_j(\cdot)$ the MSE in forecasting of forecasts i and j respectively. For any two generic forecasts i and j, we use $\text{MSE}_i(\cdot) - \text{MSE}_j(\cdot)$ as a measure of distance between these two forecasts. For N forecasts, we can choose one of them to be the benchmark, say, the first one, computing $\text{MSE}_i(\cdot) - \text{MSE}_1(\cdot)$ for $i = 2, 3, \cdots, N$. With this measure of spatial dependence at hand, we can construct a two-dimensional estimator of the asymptotic variance of $\overline{B}$ and $\widehat{B}$ following Conley(1999, Sections 3 and 4). We label $\overline{V}$ and $\widehat{V}$ the estimates of the asymptotic variances of $\overline{B}$ and of $\widehat{B}$, respectively.

Once we have estimated the asymptotic covariance of $\overline{B}$, we can test the null hypothesis $H_0 : B = 0$, by using the following t-ratio statistic:

$$t = \frac{\overline{B}}{\sqrt{\overline{V}}}.$$

By the central limit theorem, $t \xrightarrow[N\to\infty]{d} N(0,1)$ under $H_0 : B = 0$. Now consider $\widehat{t} = \frac{\widehat{B}}{\sqrt{\widehat{V}}}$, where $\widehat{V}$ is computed using $\widehat{k} = (\widehat{k}_1, \widehat{k}_2, ..., \widehat{k}_N)'$ in place of $k = (k_1, k_2, ..., k_N)'$. We have proved that $\widehat{k}_i \xrightarrow{p} k_i$ as $T \to \infty$, then the test statistics t and $\widehat{t}$ are asymptotically equivalent and therefore

$$\widehat{t} = \frac{\widehat{B}}{\sqrt{\widehat{V}}} \quad \xrightarrow[(T,N\to\infty)_{\text{seq}}]{d} \mathcal{N}(0,1).$$

$\blacksquare$

3 Monte-Carlo Study

3.1 Experiment design

We follow the setup presented in the theoretical part of this paper in which each forecast is the conditional expectation plus an additive bias term. Our DGP is a simple stationary $AR(1)$ process as described below:

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 y_{t-1} + \xi_t,\ t = 1, ...T_1, ..., T_2, ..., T & (19) \\ \xi_t &\sim \text{i.i.d.}N(0,1),\ \alpha_0 = 0, \text{ and } \alpha_1 = 0.5, & (20) \end{aligned}$$

where ξ_t is an (unpredictable) aggregate zero mean shock. We focus on one-step-ahead forecasts for simplicity. The conditional expectation of y_t is $\mathbb{E}_{t-1}(y_t) = \alpha_0 + \alpha_1 y_{t-1}$. Since ξ_t is unpredictable, the forecaster should be held accountable for $f_{i,t} - \mathbb{E}_{t-1}(y_t)$. These deviations have two terms: the individual specific biases (k_i) and the idiosyncratic or group error term ($\varepsilon_{i,t}$). Because $\xi_t \sim$i.i.d.$N(0,1)$, the optimal MSE is unity in this exercise.

The conditional expectation $\mathbb{E}_{t-1}(y_t) = \alpha_0 + \alpha_1 y_{t-1}$ is estimated using a sample of size 200, i.e., $T_1 = 200$, so that $\widehat{\alpha}_0 \simeq \alpha_0$ and $\widehat{\alpha}_1 \simeq \alpha_1$. In practice, however, forecasters may have economic incentives to make biased forecasts, and there may be other sources of mispecification arising from mispecification errors. Therefore, we assume that:

$$\begin{aligned} f_{i,t} &= \widehat{\alpha}_0 + \widehat{\alpha}_1 y_{t-1} + k_i + \varepsilon_{i,t}, & (21) \\ &= (\widehat{\alpha}_0 + k_i) + \widehat{\alpha}_1 y_{t-1} + \varepsilon_{i,t} \text{ for } t = T_1 + 1, \cdots, T,\ i = 1, ...N, \end{aligned}$$

where, $k_i = \beta k_{i-1} + u_i$, $u_i \sim$i.i.d. Uniform(a, b), $0 < \beta < 1$, and $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, ... \varepsilon_{N,t})'$, $N \times 1$, is drawn from a multivariate Normal distribution with size $R + P = T - T_1$, whose mean vector equals to zero and covariance matrix equals $\Omega = (\sigma_{ij})$. We introduce heterogeneity and spatial dependence in the distribution of $\varepsilon_{i,t}$. The diagonal elements of $\Omega = (\sigma_{ij})$ obey: $1 < \sigma_{ii} < \sqrt{10}$, and off-diagonal elements obey: $\sigma_{ij} = 0.5$ if $|i - j| = 1$, $\sigma_{ij} = 0.25$ if $|i - j| = 2$ and $\sigma_{ij} = 0$ if $|i - j| > 2$. The exact value of σ_{ii} is randomly determined through an once-and-for-all draw from a uniform random variable of size N, that is, σ_{ii} ~i.i.d. Uniform$(1, \sqrt{10})$[10].

In Equation (21), we built spatial dependence in the bias term k_i. The cross-sectional average of k_i is $\frac{a+b}{2(1-\beta)}$. The additive bias k_i is explicit in (21). It could be implicit if we had considered a structural break in the target variable as did Hendry and Clements (2002). There, an intercept change in y_t takes place right after the estimation of econometric models, biasing all forecasts. Hence, intercept correction is equivalent to bias correction. Notice that this is the case here too. A structural break as described would violate weak stationarity and that is why it is not attempted here. We set the degree of spatial dependence in k_i by letting $\beta = 0.5$. For the support of u_i, we considered two cases: (i) $a = 0$ and $b = 0.5$ and; (ii) $a = -0.5$ and $b = 0.5$. This implies that the average bias is $B = 0.5$ in (i), whereas $B = 0$ in (ii).Finally, notice that the specification of $\varepsilon_{i,t}$ satisfies Assumption 4 in Section 2, as we let $N \to \infty$.

Equation (21) is used to generate three panels of forecasts. They differ from each other in terms of the number of forecasters (N): $N = 10, 20, 40$. We assume that they all have the same number of training-sample and out-of-sample observations: $R = 50$, and $P = 50$, respectively.

In each experiment, we conduct $50,000$ simulations of data set for the three panels above. The total sample equals 300 observations ($T_1 = 200$, $R = 50$, and $P = 50$) in each panel.

3.2 Forecast approaches

In our simulations, we evaluate three forecasting methods: the feasible bias-corrected average forecast (BCAF), the weighted forecast combination, and the simple average of forecasts. Our results include aspects of the whole distribution of the bias and the MSE of computed for these methods.

For the BCAF, we use the training-sample observations to estimate $\widehat{k}_i =$

[10]The covariance matrix Ω_i does not change over simulations.

$\frac{1}{R}\sum_{t=T_1+1}^{T_2}(y_t - f_{i,t})$ and $\widehat{B} = \frac{1}{N}\sum_{i=1}^{N}\widehat{k}_i$. Then, we compute the out-of-sample forecasts $\widehat{f}_{i,t}^{BCAF} = \frac{1}{N}\sum_{i=1}^{N} f_{i,t} - \widehat{B}$, $t = T_2 + 1, ..., T$, and its respective MSE $MSE_{BCAF} = \frac{1}{P}\sum_{t=T_2+1}^{T}\left(y_t - \widehat{f}_{i,t}^{BCAF}\right)^2$.

For the weighted average forecast, we use R observations of the training sample to estimate weights (ω_i) by OLS in:

$$y = \delta\mathbf{i} + \omega_1 f_1 + \omega_2 f_2 + ... + \omega_N f_N + \varepsilon,$$

where the restriction $\sum_{i=1}^{N}\omega_i = 1$ is imposed in estimation, and $\mathbf{i}$ is an R-dimensional vector of ones, and y, f_1, $f_2, ...f_N$ denote, respectively, the $R \times 1$ vector of observations of the target variable and the N individual forecasts. The weighted forecast $\widehat{f}^{weighted} = \widehat{\delta}\mathbf{i} + \widehat{\omega}_1 f_1 + \widehat{\omega}_2 f_2 + ... + \widehat{\omega}_N f_N$ is unbiased, but its variance is grows with N for fixed R. The intercept δ plays the role of bias correction. After computing $\widehat{\delta}$ and $\{\widehat{\omega}_i\}_{i=1}^{N}$, we employ the last P observations to compute its respective MSE $MSE_{Weighted} = \frac{1}{P}\sum_{t=T_2+1}^{T}\left(y_t - \widehat{f}_{i,t}^{weighted}\right)^2$.

The last approach is the average forecast.There is no parameter to be estimated using training sample observations. Therefore, out-of-sample forecasts are computed according to the simple average $f_{i,t}^{average} = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}$, $t = T2 + 1, ..., T$, and its MSE is computed as $MSE_{Average} = \frac{1}{P}\sum_{t=T_2+1}^{T}\left(y_t - \widehat{f}_{i,t}^{average}\right)^2$.

Finally, for each approach, we also computed the out-of-sample bias. In theory, the weighted forecast and the BCAF should have out-of-sample bias close to zero, whereas it should be close to $B = \frac{a+b}{2(1-\beta)}$ for the average forecast.

3.3 Simulation Results

With the results of the 50,000 replications, we describe the empirical distributions of the bias and the MSE of all three forecasting methods. For each distribution we compute the following statistics: (i) kurtosis; (ii) skewness,

(iii) τ-th unconditional quantile, with $\tau = 0.01, 0.25, 0.50, 0.75$, and 0.99. In doing so, we seek to have a general description of all three forecasting approaches.

The main results are presented in Tables 1 and 2. In Table 1, $B = 0.5$ and in Table 2 $B = 0$. In Table 1, the average bias across simulations of the BCAF and the weighted forecast combination are practically zero. The bias of the simple average forecast is between 0.39 and 0.46, depending on N. In terms of MSE, the BCAF performs very well compared to the other two methods. The simple average has an mean MSE at least 8.7% higher, reaching 17.8% higher when $N = 40$. The weighted combination has an mean MSE at least 22.7% higher, reaching 431.3% higher when $N = 40$. This last result is a consequence of the increase in variance when we increase N, with R fixed, and N/R close to unity. Notice that, when $N = 40$ $N/R = 0.8$. As stressed above, this results is expected and it would not happen if N/R was small. Since $R = 50$, increasing N from 10 to 40 reveals the curse-of-dimensionality of the weighted forecast combination. For the other two methods, the distribution of MSE shrinks with N. For the BCAF, we reach an average MSE of 1.147 when $N = 40$, whereas the theoretical optimal MSE is 1.000.

Table 2 presents the results when $B = 0$. In this case, the optimal forecast is the simple average, since there is no need to estimate a bias-correction term. In terms of MSE, comparing the simple-average forecast with the BCAF, we observe that they are almost identical – the mean MSE of the BCAF is about 1% higher than that of the average forecast, showing that not much is lost in terms of MSE when we perform an unwanted bias correction. Bias is also unaffected by the correction. The behavior of the weighted average forecast is identical to that in Table 1.

4 Empirical Application

Professional forecasts guide market participants and inform them about future economic conditions. However, many analysts argue that forecasters might strategically bias forecasts as long as they receive economic incentives to do so. The importance of microeconomic incentives for forecasters and analysts is stressed by a number of empirical studies, such as Ehrbeck and Waldmann (1996), Graham (1999), Hong et al. (2000), Lamont (2002), Welch (2000), and Zitzewitz (2001).

In this section, we present an application of the method proposed here for the case of forecast surveys, showing that bias correction can indeed help in forecasting. We also test the hypothesis that forecasters behave strategically using a dual loss function. When this fact is accounted for, the bias-corrected average forecast introduced in this paper outperforms simple forecast averages (consensus). It is important to stress that, although our techniques were conceived for a large N, T environment, the empirical results here show the usefulness of our method even in a small N, T environment. Also, the forecasting gains from bias correction are non-trivial.

4.1 The Central Bank of Brazil's "Focus Forecast Survey"

The "Focus Forecast Survey," collected by the Central Bank of Brazil, is a unique panel database of forecasts. It contains forecast information on almost 120 institutions, including commercial banks, asset-management firms, and non-financial institutions, which are followed throughout time with a reasonable turnover. Forecasts have been collected since 1998, on a monthly frequency, and a fixed horizon, which potentially can serve to approximate a large N, T environment for techniques designed to deal with unbalanced panels – which is not the case studied here. Besides the large size of N and T in the Focus Survey, it also has the following desirable features: the anonymity of forecasters is preserved, although the names of the top-five forecasters for a given economic variable is released by the Central Bank of Brazil; forecasts are collected at different frequencies (monthly, semi-annual, annual), as well as at different forecast horizons (e.g., short-run forecasts are obtained for h from 1 to 12 months); there is a large array of macroeconomic time series included in the survey.

To save space, below we focus our analysis on the behavior of forecasts of the monthly inflation rate in Brazil (π_t), in percentage points, as measured by the official Consumer Price Index (CPI), computed by FIBGE. In order to obtain the largest possible balanced panel ($N \times T$), we used $N = 18$ and a time-series sample period covering the period 2002:11 through 2006:3 ($T = 41$). Of course, in the case of a survey panel, there is no estimation sample. We chose the first 26 time observations to compute $\widehat{B}$ – the average bias – leaving 18 time-series observations for out-of-sample forecast evaluation. The forecast horizon chosen was $h = 6$, this being an important horizon to determine future monetary policy within the Brazilian Inflation-Targeting program.

The results of our empirical exercise are presented in Tables 3 and 4. They show that the average bias is positive for the 6-month horizon – 0.06187 – and marginally significant, with a p-value of 0.063. This is a sizable bias – approximately 0.745 percentage points in a yearly basis, for an average inflation rate of 5.266% a year. Out-of-sample forecast comparisons between the simple average and the bias-corrected average forecast show that the former has an MSE 18.2% bigger than that of the latter. We also computed the MSE of the weighted forecast. Since we have $N = 18$ and $R = 26$, $N/R = 0.69$, and our estimate could not avoid the curse of dimensionality yielding a MSE 390.2% bigger than that of the BCAF.

5 Conclusions and Extensions

In this paper, we propose a novel approach to econometric forecasting of stationary and ergodic series within a panel-data framework, where the number of forecasts and the number of time periods increase without bounds. The advantages of our approach are many. First, only in this context we can fully understand why the pooling of forecasts works in practice under an MSE loss function. Second, we can also propose improvements on simple forecast-combination schemes. Here we propose the bias-corrected average forecast. Third, the techniques discussed here are applicable in two important contexts: when forecasts are a result of model estimation, and when they are the result of opinion polls.

The basis of our method is to decompose individual forecasts into three components: the series being forecast, a time-invariant forecast bias, and a zero-mean forecast error. The series being forecast is viewed as a *common feature* of all individual forecasts. Standard tools from panel-data asymptotic theory are used to devise an optimal forecasting combination that has a zero limiting mean-squared forecast error. This optimal forecast combination uses equal weights and a bias-correction term. The use of equal weights avoids estimating forecast weights, which contributes to reduce forecast variance, although potentially at the cost of an increase in bias. The use of a bias-correction term eliminates any possible detrimental effect arising from equal weighting. We label this optimal forecast as the *bias-corrected average forecast*.

In theory – large N and T – the use of a bias-corrected average forecast is potentially superior to the use of any single forecast and is equal or superior to any other combining technique. Moreover, in practice – small N and/or T – an important element of the use of the bias-corrected average forecast is that the forecast combination puzzle works to our advantage, now augmented with a bias-correction term. Hence, there will be situations in which we can improve upon the simple average forecast by using bias-correction, and others which we cannot. Our framework offers a statistical test for excluding the bias-correction term.

The Monte-Carlo experiment and the empirical analyses performed here show the usefulness of our new approach. Regarding model misspecification bias, the Monte-Carlo experiment shows important improvement over conventional combination techniques. In the empirical exercise, we showed that using our method leads to an improvement in forecasting accuracy under MSE loss – from about 10% to about 60%. As one should expect, higher gains for bias correction are observed when the null hypothesis of a zero bias is rejected in testing.

For reasons of space, we refrain from fully discussing here natural extensions of our proposed method. A partial account of those includes the following:

1. In the panel of forecasts:

$$\left(f_{i,t}^h - y_t\right) = k_i + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2, \tag{22}$$

   we impose a unity coefficient for y_t, but we could have had an encompassing panel-regression system:

$$f_{i,t}^h = \beta_i y_t + k_i + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2, \tag{23}$$

   where β_i can be interpreted as the *beta* of forecast-model i vis-à-vis y_t. A natural hypothesis to test is $H_0 : \beta_i = 1$, for all i, which can be implemented using standard panel techniques.

2. There may be instances where forecast models produce forecasts that are too highly correlated. In theory, this may prevent a weak law-of-large-numbers from holding for the error terms. In this case we can combine pooling of information and pooling of forecasts:

$$\left(f_{i,t}^h - y_t\right) = k_i + \sum_{k=1}^{K} \beta_{i,k} f_{k,t} + \eta_{i,t}, \qquad i = 1, 2, \ldots, N, \qquad T_1 < t < T_2, \tag{24}$$

   where $f_{k,t}$ are zero-mean pervasive factors and, as is usual in factor analysis,

$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\eta_{i,t} = 0$. In this context:

$$\varepsilon_{i,t} = \sum_{k=1}^{K}\beta_{i,k}f_{k,t} + \eta_{i,t}, \qquad i = 1,2,\ldots,N, \qquad T_1 < t < T_2.$$

Thus, factor and principal-component analyses (Stock and Watson(1999 and 2002a and b) and Forni et al. (2000, 2003)) are combined with the idea of bias-corrected average forecasts. In that sense, we combine *pooling of forecasts* with *pooling of information* in the same model.

3. The final extension considered here is to allow for a time-varying bias term ξ_t. In this case,

$$\left(f_{i,t}^{h} - y_t\right) = k_i + \xi_t + \varepsilon_{i,t}, \qquad i = 1,2,\ldots,N, \qquad T_1 < t < T_2. \quad (25)$$

The techniques of Fuller and Battese (1974) can be a starting point to generate consistent estimates of k_i and ξ_t in a context where N and T are large.

References

[1] Araujo, F., Issler, J.V. and Fernandes, M.(2006), "A Stochastic Discount Factor Approach to Asset Pricing Using Panel Data," Working Paper: EPGE-FGV # 628, downloadable from http://epge.fgv.br/portal/arquivo/2155.pdf

[2] Bai, J., (2005), "Panel Data Models with Interactive Fixed Effects," Working Paper: New York University.

[3] Bai, J., and S. Ng, (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.

[4] Bai, J. and S. Ng, (2004), "Evaluating Latent and Observed Factors in Macroeconomics and Finance," Working Paper: University of Michigan.

[5] Beran, Jan, (1998), "*Statistics for Long-Memory Processes.*" Chapman and Hall – CRC.

[6] Bollerslev, T., (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, Vol.31, pp.307-327.

[7] Chamberlain, Gary, and Rothschild, Michael, (1983). "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, vol. 51(5), pp. 1281-1304.

[8] Clements, M.P. and D.F. Hendry, 1998, *Forecasting Economic Time Series*, Cambridge University Press.

[9] Clements, M.P. and D.F. Hendry, 2006, Forecasting with Breaks in Data Processes, in C.W.J. Granger, G. Elliott and A. Timmermann (eds.) *Handbook of Economic Forecasting*, pp. 605-657, Amsterdam, North-Holland.

[10] Conley, T.G., 1999, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, Vol. 92 Issue 1, pp. 1-45.

[11] Connor, G., and R. Korajzcyk (1986), "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373-394.

[12] Connor, G. and Korajczyk, R. (1993), "A test for the number of factors in am approximate factor structure," *Journal of Finance* 48, 1263 - 1291.

[13] Ehrbeck, T., Waldman, R., 1996. Why are professional forecasters biased? Agency versus behavioral explanations. *Quarterly Journal of Economics* 111, 21-40.

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

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

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

[17] Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, pp. 987-1006.

[18] Engle, R. F., Issler, J. V., 1995, "Estimating common sectoral cycles," *Journal of Monetary Economics*, vol. 35, 83–113.

[19] Engle, R.F. and Kozicki, S. (1993). "Testing for Common Features", *Journal of Business and Economic Statistics*, 11(4): 369-80.

[20] Forni, M., Hallim, M., Lippi, M. and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation", *Review of Economics and Statistics*, 2000, vol. 82, issue 4, pp. 540-554.

[21] Forni M., Hallim M., Lippi M. and Reichlin L., 2003 "The Generalized Dynamic Factor Model one-sided estimation and forecasting," *Journal of the American Statistical Association*, forthcoming.

[22] Forni, M., M. Hallim, M. Lippi and L. Reichlin (2004), "The generalized factor model: consistency and rates", *Journal of Econometrics* 119:231-255.

[23] Fuller, Wayne A. and George E. Battese, 1974, "Estimation of linear models with crossed-error structure," *Journal of Econometrics*, Vol. 2(1), pp. 67-78.

[24] Giacomini, Raffaella and Halbert White, 2006, "Tests of Conditional Predictive Ability," *Econometrica*, vol. 74(6), pp. 1545-1578.

[25] Graham, J., 1999. Herding among investment newsletters: theory and evidence. Journal of Finance 54, 231-268.

[26] Granger, C.W.J., and R.Ramanathan (1984), “Improved methods of combining forecasting”, *Journal of Forecasting* 3:197–204.

[27] Hendry, D.F. and M.P. Clements (2002), “Pooling of forecasts”, *Econometrics Journal*, 5:1-26.

[28] Hong, H., Kubik, J.D., Solomon, A., 2000. Security analysts ´career concerns and herding of earning forecasts. *RAND Journal of Economics* 31, 121-144.

[29] Hoogstrate, Andre J., Franz C. Palm, Gerard A. Pfann, 2000, “Pooling in Dynamic Panel-Data Models: An Application to Forecasting GDP Growth Rates,” *Journal of Business and Economic Statistics*, Vol. 18, No. 3, pp. 274-283.

[30] Issler, J. V., Vahid, F., 2001, “Common cycles and the importance of transitory shocks to macroeconomic aggregates,” *Journal of Monetary Economics*, vol. 47, 449–475.

[31] Issler, J. V., Vahid, F., 2006, “The missing link: Using the NBER recession indicator to construct coincident and leading indices of economic activity,” *Annals Issue of the Journal of Econometrics* on *Common Features*, vol. 132(1), pp. 281-303.

[32] Kang, H. (1986), "Unstable Weights in the Combination of Forecasts," *Management Science* 32, 683-95.

[33] Lamont, O., 2002. Macroeconomic forecasts and microeconomic forecasters. *Quarterly Journal of Economics* 114, 293-318.

[34] Lima, L.R. and Z. Xiao, 2007. ”Do Shocks Last Forever? Local Persistency in Economic Time Series”. *Journal of Macroeconomics, vol.29, n. 2.*

[35] Palm, Franz C. and Arnold Zellner, 1992, "To combine or not to combine? issues of combining forecasts," *Journal of Forecasting*, Volume 11, Issue 8 , pp. 687-701.

[36] Pesaran, M.H., (2005), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." Working Paper: Cambridge University, forthcoming in *Econometrica*.

[37] Phillips, P.C.B., 1995, "Lecture Notes," downloadable from the author's webpage: Yale University.

[38] Phillips, P.C.B. and H.R. Moon, 1999, "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, vol. 67 (5), pp. 1057–1111.

[39] Phillips, P.C.B., H.R. Moon and Z. Xiao, 2001, "How to estimate autoregressive roots near unity," *Econometric Theory* 17, 29-69.

[40] Ross, S.A. (1976), "The arbitrage theory of capital asset pricing", *Journal of Economic Theory*, 13, pp. 341-360.

[41] Smith, Jeremy and Kenneth F. Wallis, 2005, "Combining Point Forecasts: The Simple Average Rules, OK?" Working Paper: Department of Economics, University of Warwick.

[42] Stock, J. and Watson, M., "Forecasting Inflation", *Journal of Monetary Economics*, 1999, Vol. 44, no. 2.

[43] Stock, J. and Watson, M., "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, April 2002a, Vol. 20 No. 2, 147-162.

[44] Stock, J. and Watson, M., "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 2002b.

[45] Stock, J. and Watson, M., 2006, "Forecasting with Many Predictors,"In: Elliott, G., C.W.J. Granger, and A. Timmermann, 2006, Editors, *Handbook of Economic Forecasting*, Amsterdam: North-Holland, Chapter 10, pp. 515-554.

[46] Timmermann, A., 2006, "Forecast Combinations," In: Elliott, G., C.W.J. Granger, and A. Timmermann, 2006, Editors, *Handbook of Economic Forecasting*, Amsterdam: North-Holland, Chapter 4, pp. 135-196.

[47] Vahid, F. and Engle, R. F., 1993, "Common trends and common cycles," *Journal of Applied Econometrics*, vol. 8, 341–360.

[48] Vahid, F., Engle, R. F., 1997, "Codependent cycles," *Journal of Econometrics*, vol. 80, 199–221.

[49] 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.

[50] Welch, I., 2000. Herding among security analysts. *Journal of Financial Economics* 58, 369-396.

[51] Zitzewitz, E., 2001. Measuring herding and exaggeration by equity analysts. Unpublished working paper. Graduate School of Business, Stanford University.

A Tables and Figures

Table 1: Monte-Carlo Results

$T_2 = 50$ $a = 0; b = 0.5$

	Bias Distributions			MSE Distributions		
	BCAF	Average	Weighted	BCAF	Average	Weighted
				$N = 10$		
skewness	-0.009	0.003	0.004	0.404	0.438	0.632
kurtosis	3.043	3.037	3.038	3.221	3.298	3.823
mean	0.000	0.391	-0.001	1.561	1.697	1.916
$\tau = 0.01$	-0.590	-0.091	-0.643	0.911	0.989	1.080
$\tau = 0.25$	-0.169	0.252	-0.186	1.337	1.448	1.608
$\tau = 0.50$	-0.001	0.391	-0.001	1.540	1.672	1.872
$\tau = 0.75$	0.167	0.530	0.184	1.763	1.918	2.179
$\tau = 0.99$	0.578	0.879	0.641	2.394	2.642	3.138
				$N = 20$		
skewness	0.010	-0.013	0.011	0.442	0.444	0.961
kurtosis	3.115	3.084	3.138	3.321	3.321	5.011
mean	0.000	0.440	-0.002	1.286	1.466	2.128
$\tau = 0.01$	-0.532	-0.001	-0.690	0.754	0.853	1.117
$\tau = 0.25$	-0.153	0.316	-0.195	1.098	1.247	1.723
$\tau = 0.50$	-0.001	0.440	-0.004	1.266	1.444	2.053
$\tau = 0.75$	0.151	0.565	0.192	1.452	1.659	2.448
$\tau = 0.99$	0.535	0.876	0.687	1.987	2.275	3.851
				$N = 40$		
skewness	-0.015	-0.006	-0.016	0.438	0.448	2.852
kurtosis	3.147	3.090	3.600	3.324	3.338	22.203
mean	0.000	0.465	0.000	1.147	1.351	6.094
$\tau = 0.01$	-0.515	0.050	-1.209	0.673	0.786	2.165
$\tau = 0.25$	-0.145	0.346	-0.315	0.980	1.150	4.021
$\tau = 0.50$	0.000	0.465	0.002	1.130	1.331	5.337
$\tau = 0.75$	0.141	0.583	0.312	1.295	1.529	7.243
$\tau = 0.99$	0.509	0.876	1.209	1.771	2.100	17.669

Table 2: Monte-Carlo Results

$T_2 = 50,\ a = -0.5; b = 0.5$

	Bias Distributions			MSE Distributions		
	BCAF	Average	Weighted	BCAF	Average	Weighted
				$N = 10$		
skewness	-0.009	0.005	0.004	0.404	0.395	0.632
kurtosis	3.043	3.016	3.038	3.221	3.228	3.823
mean	0.000	0.000	-0.001	1.561	1.547	1.916
$\tau = 0.01$	-0.590	-0.511	-0.643	0.911	0.905	1.080
$\tau = 0.25$	-0.169	-0.149	-0.186	1.337	1.326	1.608
$\tau = 0.50$	-0.001	0.000	-0.001	1.540	1.526	1.872
$\tau = 0.75$	0.167	0.147	0.184	1.763	1.745	2.179
$\tau = 0.99$	0.578	0.516	0.641	2.394	2.369	3.138
				$N = 20$		
skewness	0.010	-0.015	0.011	0.442	0.414	0.961
kurtosis	3.115	3.071	3.138	3.321	3.283	5.011
mean	0.000	0.000	-0.002	1.286	1.272	2.128
$\tau = 0.01$	-0.532	-0.462	-0.690	0.754	0.746	1.117
$\tau = 0.25$	-0.153	-0.130	-0.195	1.098	1.089	1.723
$\tau = 0.50$	-0.001	0.000	-0.004	1.266	1.254	2.053
$\tau = 0.75$	0.151	0.130	0.192	1.452	1.435	2.448
$\tau = 0.99$	0.535	0.456	0.687	1.987	1.951	3.851
				$N = 40$		
skewness	-0.015	-0.005	-0.016	0.438	0.414	2.852
kurtosis	3.147	3.090	3.600	3.324	3.266	22.203
mean	-0.002	0.000	0.000	1.147	1.133	6.094
$\tau = 0.01$	-0.515	-0.426	-1.209	0.673	0.667	2.165
$\tau = 0.25$	-0.145	-0.123	-0.315	0.980	0.971	4.021
$\tau = 0.50$	-0.002	0.000	0.002	1.130	1.116	5.337
$\tau = 0.75$	0.141	0.121	0.312	1.295	1.278	7.243
$\tau = 0.99$	0.509	0.424	1.209	1.771	1.733	17.669

Notes: ????(A) $RMSE_i^p$ $i = 1, 2$ denotes the mean of the relative MSE of: (1) the simple average forecast ($RMSE_1^p$), and, (2) the weighted-average forecast ($RMSE_2^p$). In both cases, the MSE of the bias-corrected average forecast is taken as *numeraire*. (B) The superscript p indicates the number of observations in the training sample.

Table 3: The Brazilian Central Bank Focus Survey
Computing Average Bias and Testing the No-Bias Hypothesis

Horizon (h)	Avg. Bias $\widehat{B}$	$H_0 : B = 0$ p-value
6	0.06187	0.063

Notes: (1) $N = 18$, $R = 26$, $P = 15$, and $h = 6$ months ahead.

Table 4: The Brazilian Central Bank Focus Survey
Comparing the MSE of Simple Average Forecast with that of the Bias-Corrected Average Forecast and the Weighted Average Forecast

Forecast Horizon (h)	(a) MSE BCAF	(b) MSE Avg. Forecast	(c) MSE Weighted Avg. Forecast	(b)/(a)	(c)/(a)
6	0.0683	0.0808	0.2665	1.182	3.902

Notes: (1) $N = 18$, $R = 23$ and $P = 18$, and $h = 6$ months ahead.