



FUNDAÇÃO
GETULIO VARGAS

EPGE

Escola de Pós-Graduação
em Economia

Ensaio Econômico

Escola de

Pós Graduação

em Economia

da Fundação

Getúlio Vargas

Nº 650

ISSN 0104-8910

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

João Victor Issler, Luiz Renato Lima

Setembro de 2007

**Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões
neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação
Getulio Vargas.**

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

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

First Draft: December, 2006.

Revised: September, 2007.

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 (feasible) bias-corrected average forecast. Using panel-data sequential asymptotics we show that it is potentially superior to other techniques in several contexts. In particular, it is asymptotically equivalent to the conditional expectation, i.e., has an optimal limiting mean-squared error. We also develop a zero-mean test for the average bias and discuss the forecast-combination puzzle in small and large samples. Monte-Carlo simulations are conducted to evaluate the performance of the feasible bias-corrected average forecast in finite samples. An empirical exercise, based upon data

*We are especially grateful for the comments and suggestions given by two anonymous referees, Marcelo Fernandes, Wagner Gaglianone, Antonio Galvão, Raffaella Giacomini, Clive Granger, Roger Koenker, 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 excellent research assistance and gratefully acknowledge the 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 written. The usual disclaimer applies.

[†]Corresponding Author.

from a well known survey is also presented. Overall, these results show promise for the feasible bias-corrected average forecast.

Keywords: Forecast Combination, Forecast-Combination Puzzle, Common Features, Panel-Data, Bias-Corrected Average Forecast.

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 in the mean-squared error (MSE) sense.

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. This argument was considered in Palm and Zellner (1992) who asked the question “*to pool or not to pool*” forecasts? Recently, Timmermann (2006) used risk diversification – a principle so keen in finance – to defend pooling of forecasts. Of course, to obtain this WLLN result, at least the number of forecasts has to diverge ($N \rightarrow \infty$), 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, with $N, T \rightarrow \infty$, is that forecasting is frequently thought to be a time-series experiment, not a panel-data experiment.

In this paper, we propose a novel approach to econometric forecast of stationary and ergodic series within a panel-data framework. First, we use a two-way decomposition for the forecast error (Wallace and Hussein (1969)), where individual errors are the sum of a time-invariant forecast bias, an unforecastable aggregate zero-mean shock, and an idiosyncratic (or subgroup) zero-mean error term. Second, we show the equivalence between this two-way decomposition and a model where forecasts are a biased and error-ridden version of the optimal forecast in the MSE sense – the conditional

expectation of the series being forecast. Indeed, the latter is the *common feature* of all individual forecasts (Engle and Kozicki (1993)), while individual forecasts deviate from the optimal because of forecast misspecification; see the reasons listed in Palm and Zellner. Third, when $N, T \rightarrow \infty$, and we use standard tools from panel-data asymptotic theory, we show that the pooling of forecasts delivers optimal limiting forecasts in the MSE sense. In our key result, we prove that, in the limit, the feasible *bias-corrected average forecast* – equal weights in combining forecasts coupled with an estimated bias-correction term – is an optimal forecast identical to the conditional expectation.

The feasible bias-corrected average forecast is also parsimonious besides being optimal. The only parameter we need to estimate is the mean bias, for which we show consistency under 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.

As a by-product of the use of panel-data asymptotic methods, with $N, T \rightarrow \infty$, we advanced the understanding of the forecast combination puzzle. The key issue is that simple averaging requires no estimation of weights, while optimal weights requires estimating N weights that grow unbounded in the asymptotic setup. We show that there is no puzzle under certain asymptotic paths for N and T , but not for all. We fully characterize them here. We are also able to discuss the puzzle in small samples, linking its presence to the curse of dimensionality which plagues so many estimators throughout econometrics¹.

Despite the scarcity of panel-data studies on the pooling of forecasts², there has been panel-data research on forecast focusing on the pooling of information; see Stock and Watson (1999 and 2002a and b) and Forni et al. (2000, 2003). Pooling forecasts is related to forecast combination and

¹We thank Roger Koenker for suggesting this asymptotic exercise to us, and an anonymous referee for casting the puzzle in terms of the curse of dimensionality.

²The notable exception is Palm and Zellner (1992), who discuss “to pool or not to pool” forecasts using a two-way decomposition. They make very limited use of the panel dimension of forecasts in their discussion. Davies and Lahiri (1995) use a three-way decomposition, but focus on forecast rationality instead of combination.

operates a reduction on the space of forecasts. Pooling information operates a reduction on a set of highly correlated regressors. Forecasting can benefit from the use of both procedures, since, in principle, both yield asymptotically optimal forecasts in the MSE sense.

A potential limitation on the literature on pooling of information is that pooling is performed in a linear setup, and the statistical techniques employed were conceived as highly parametric – principal-component and factor analysis. That is a problem if the conditional expectation is not a linear function of the conditioning set or if the parametric restrictions used (if any) are too stringent to fit the information being pooled. In this case, pooling forecasts will be a superior choice, since the forecasts being pooled need not be the result of estimating a linear model under a highly restrictive parameterization. On the contrary, these models may be non-linear, non-parametric, and even unknown to the econometrician, as is the case of using a survey of forecasts. Moreover, the components of the two-way decomposition employed here are estimated using non-parametric techniques, dispensing any distributional assumptions. This widens the application of the methods discussed in this paper.

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 and intercept correction, perhaps best represented by the work of Bates and Granger (1969), Granger and Ramanathan (1984), Palm and Zellner (1992), Forni et al. (2000, 2003), Hendry and Clements (2002), Stock and Watson (2002a and b), Elliott and Timmermann (2003, 2004, 2005), Hendry and Mizon (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 and risk diversification 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 Chamberlain and Rothschild (1983), Connor and

Korajczyk (1986), Phillips and Moon (1999), Bai and Ng (2002), Bai (2005), and Pesaran (2005). Indeed, our approach borrows from finance the idea that we can only diversify idiosyncratic risk but not systematic risk. The latter is associated with the common element of all forecasts – the conditional expectation term – which is to what a specially designed forecast average converges to.

The rest of the paper is divided as follows. Section 2 presents our main results and the assumptions needed to derive them. 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.

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 $f_{i,t}^h$, $i = 1, 2, \dots, N$. 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 .

We consider 3 consecutive distinct time sub-periods, where time is indexed by $t = 1, 2, \dots, T_1, \dots, T_2, \dots, T$. The first sub-period E is labeled the “estimation sample,” where models are usually fitted to forecast y_t in the subsequent period, if that is the case. The number of observations in it is $E = T_1 = \kappa_1 \cdot T$, comprising $(t = 1, 2, \dots, T_1)$. For the other two, we follow the standard notation in West (1996). The sub-period R (for regression) is

labeled the post-model-estimation or “training sample”, where realizations of y_t are usually confronted with forecasts produced in the estimation sample, and weights and bias-correction terms are estimated, if that is the case. It has $R = T_2 - T_1 = \kappa_2 \cdot T$ observations in it, comprising $(t = T_1 + 1, \dots, T_2)$. The final sub-period is P (for prediction), where genuine out-of-sample forecast is entertained. It has $P = T - T_2 = \kappa_3 \cdot T$ observations in it, comprising $(t = T_2 + 1, \dots, T)$. Notice that $0 < \kappa_1, \kappa_2, \kappa_3 < 1$, $\kappa_1 + \kappa_2 + \kappa_3 = 1$, and that the number of observations in these three sub-periods keep a fixed proportion with T – respectively, κ_1 , κ_2 and κ_3 – being all $O(T)$. This is an important ingredient in our asymptotic results for $T \rightarrow \infty$.

We now compare our time setup with that of West. He only considers two consecutive periods: R data points are used to estimate models and the subsequent P data points are used for prediction. His setup does not require estimating bias-correction terms or combination weights, so there is no need for an additional sub-period for estimating the models that generate the $f_{i,t}^h$ 's³. In the case of surveys, since we do not have to estimate models, our setup is equivalent to West's. Indeed, in his setup, $R, P \rightarrow \infty$ as $T \rightarrow \infty$, and $\lim_{T \rightarrow \infty} R/P = \pi \in [0, \infty]$. Here⁴:

$$\lim_{T \rightarrow \infty} \frac{R}{P} = \frac{\kappa_2}{\kappa_3} = \pi \in (0, \infty).$$

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 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 $N \rightarrow \infty$.

³Notice that the estimated models generate the $f_{i,t}^h$'s, but model estimation, bias-correction estimation and weight estimation cannot be performed all within the same sub-sample in an out-of-sample forecasting exercise.

⁴To include the supports of $\pi \in [0, \infty]$ we must, asymptotically, give up having either a training sample or a genuine out-of-sample period.

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 risk 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 $\frac{1}{N} \sum_{i=1}^N f_{i,t}^h$ usually outperforms individual forecasts $f_{i,t}^h$ shows our inability to approximate $\mathbb{E}_{t-h}(y_t)$ reasonably well with individual models. However, since $\mathbb{E}_{t-h}(y_t)$ is optimal, this is exactly what these individual models should be doing.

With this motivation, our setup writes the $f_{i,t}^h$'s as approximations to the optimal forecast as follows:

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

where k_i is the individual model time-invariant bias and $\varepsilon_{i,t}$ is the individual model error term in approximating $\mathbb{E}_{t-h}(y_t)$. Here, the optimal forecast is a common feature of all individual forecasts and k_i and $\varepsilon_{i,t}$ arise because of forecast misspecification⁵. We can always decompose the series y_t into $\mathbb{E}_{t-h}(y_t)$ and an unforecastable component ζ_t , such that $\mathbb{E}_{t-h}(\zeta_t) = 0$ in:

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

Combining (1) and (2) yields,

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

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

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

⁵If 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.

It has been largely used in econometrics dating back to Wallace and Hussein (1969), Amemiya (1971), Fuller and Battese (1974) and Baltagi (1980). Palm and Zellner (1992) employ a two-way decomposition to discuss forecast combination in a Bayesian and a non-Bayesian setup⁶, and Davies and Lahiri (1995) employed a three-way decomposition to investigate forecast rationality within the “Survey of Professional Forecasts.”

By construction, our framework in (4) specifies explicit sources of forecast errors that are found in both y_t and $f_{i,t}^h$; see also the discussion in Palm and Zellner and 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 forecasts⁷. 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 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 $N, T \rightarrow \infty$ ⁸.

⁶Palm and Zellner show that the performance of non-Bayesian combinations obey the following MSE rank: (i) the unfeasible weighted forecast with known weights performs better or equal to the simple average forecast, and (ii) the simple average forecast may perform better than the feasible weighted forecast with estimated weights. Our main result is that the feasible bias-corrected average forecast is optimal under sequential asymptotics. We also propose an explanation to the forecast-combination puzzle based on the curse of dimensionality. Critical to these results is the use of large N, T asymptotic theory.

⁷Because 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.

⁸Davies and Lahiri considered a three-way decomposition with h as an added dimension. The focus of their paper is forecast rationality. In their approach, η_t and $\varepsilon_{i,t}$ depend on

From the perspective 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, 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.,

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

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 of 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

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 decomposition. 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.

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 a MSE risk function? Laster, Bennett and Geoum (1999), Patton and Timmermann (2006), and Batchelor (2007) list different arguments consistent with forecasters having a non-quadratic loss function⁹. 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 quadratic 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 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*(\cdot) is a

⁹Palm and Zellner list additional reasons for bias in forecasts: carelessness; the use of a poor or defective information set or incorrect model; and errors of measurement. Lack of information or of full knowledge of the data-generating process (DGP) of y_t will ultimately lead to forecast misspecification and bias. Under some conditions, regressors having a structural break can also lead to forecast bias for a given model while weak-stationarity of y_t is preserved.

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 . Otherwise, the order is $h-1$. 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 $\{\varepsilon_t\}$ 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

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

Non-ergodicity 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).

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 $\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |\mathbb{E}(\xi_{i,t} \xi_{j,t})| = 0$ 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:

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

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 $\epsilon_{i,t}$, $i = 1, 2, \dots, N$.*

Proof. Start with:

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

Then, the MSE of individual forecasts is:

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

where $\sigma_{\epsilon_i}^2$ is the variance of $\epsilon_{i,t}$. Assumption 1 is used in the second 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 $N \rightarrow \infty$, the mean-squared error in forecasting y_t , combining all possible individual forecasts $f_{i,t}^h$, is*

$$MSE_{average} = \mathbb{E} \left(\underset{N \rightarrow \infty}{plim} \frac{1}{N} \sum_{i=1}^N f_{i,t}^h - y_t \right)^2 = B^2 + \sigma_\eta^2.$$

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

$$\frac{1}{N} \sum_{i=1}^N f_{i,t}^h - y_t = \frac{1}{N} \sum_{i=1}^N k_i + \eta_t + \frac{1}{N} \sum_{i=1}^N \epsilon_{i,t}.$$

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

$$\underset{N \rightarrow \infty}{plim} \frac{1}{N} \sum_{i=1}^N k_i + \eta_t + \underset{N \rightarrow \infty}{plim} \frac{1}{N} \sum_{i=1}^N \epsilon_{i,t}. \quad (9)$$

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

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

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

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

where, for all i , $b_{i,0} = 1$, $\sum_{j=0}^{\infty} b_{i,j}^2 < \infty$, 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 $\xi_{i,t-k}$, $k = 1, 2, \dots$. Therefore, we need only to consider the sum of the variances of terms of the form $\frac{1}{N} \sum_{i=1}^N b_{i,k} \xi_{i,t-k}$. These variances are given by:

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

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

$$\begin{aligned} \text{VAR} \left(\frac{1}{N} \sum_{i=1}^N b_{i,k} \xi_{i,t-k} \right) &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N b_{i,k} b_{j,k} \mathbb{E} (\xi_{i,t} \xi_{j,t}) \leq \\ \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |b_{i,k} b_{j,k} \mathbb{E} (\xi_{i,t} \xi_{j,t})| &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |b_{i,k} b_{j,k}| |\mathbb{E} (\xi_{i,t} \xi_{j,t})| \leq \quad (12) \\ \left(\max_{i,j} |b_{i,k} b_{j,k}| \right) \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |\mathbb{E} (\xi_{i,t} \xi_{j,t})| &\quad (13) \end{aligned}$$

Hence:

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

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

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

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,t} = 0.$$

Therefore,

$$\begin{aligned} \mathbb{E} \left(\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f_{i,t}^h - y_t \right)^2 &= \mathbb{E} (B + \eta_t)^2 \\ &= B^2 + \sigma_{\eta}^2. \end{aligned} \quad (14)$$

■

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:

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

where $Bias_i^2 = k_i^2$ and $VAR_i = \sigma_{\eta}^2 + \sigma_{\varepsilon_i}^2$. Proposition 2 shows that averaging forecasts reduces variance, but not necessarily MSE,

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

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 (BCAF).

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 $\text{plim}_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N f_{i,t}^h - \frac{1}{N} \sum_{i=1}^N k_i \right) = y_t + \eta_t$ and has a mean-squared error as follows:*

$$MSE_{BCAF} = \mathbb{E} \left[\text{plim}_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N f_{i,t}^h - \frac{1}{N} \sum_{i=1}^N k_i \right) - y_t \right]^2 = \sigma_\eta^2. \text{ Therefore, it is an optimal forecasting device in the MSE sense.}$$

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

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

leading to:

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

■

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 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 $\lim_{N \rightarrow \infty} \sum_{i=1}^N \omega_i = 1$. Then, under Assumptions 1-4,*

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

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 cannot perform better than $1/N$ under bias correction. Therefore, there is no forecast-combination puzzle in the context of population 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 $\sum_{i=1}^N \omega_i = 1$:

$$y = \delta \mathbf{i} + \omega_1 f_1 + \omega_2 f_2 + \dots + \omega_N f_N + v, \quad (16)$$

where y denotes the $R \times 1$ vector of observations of the target variable, f_1, f_2, \dots, 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 $T_1 + 1, \dots, T_2$ (i.e., over the training sample). On the one hand, because regression (16) includes an intercept, the forecast $\widehat{f} = \widehat{\delta} \mathbf{i} + \widehat{\omega}_1 f_1 + \widehat{\omega}_2 f_2 + \dots + \widehat{\omega}_N f_N$ 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 $N, T \rightarrow \infty$. In both cases, feasible estimates require $N < R$. 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 an explanation to the forecast-combination puzzle. Asymptotically, feasibility requires:

$$0 < \lim_{N, T \rightarrow \infty} \frac{N}{R} = \lim_{N, T \rightarrow \infty} \frac{N/T}{\frac{T_2 - T_1}{T}} = \frac{\lim_{N, T \rightarrow \infty} N/T}{\kappa_2} < 1, \quad (17)$$

which implies not only that $N \rightarrow \infty$ at a smaller rate than T , but that $\lim_{N, T \rightarrow \infty} N/T < \kappa_2$. Recall that $\kappa_2 = 1 - \kappa_1 - \kappa_3$, hence, $\lim_{N, T \rightarrow \infty} N/T \ll 1$. As long as this 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 trade-off 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 3. Hence, we are able to improve upon the simple average forecast¹⁰.

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 immediately 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 $f_{i,t}^h$, $i = 1 \dots N$, for the R training-sample observations. Hence, one can form a panel of forecasts:

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

where it becomes obvious that k_i represents the fixed effect of 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 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

¹⁰Only 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 $N = 3$.

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 $(T, N \rightarrow \infty)_{\text{seq}}$.

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:*

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

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, $T_1, T_2, R \rightarrow \infty$, as $T \rightarrow \infty$. Then, as $T \rightarrow \infty$,

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

Given that we observe $f_{i,t}^h$ and y_t , we propose the following consistent esti-

mator for k_i , as $T \rightarrow \infty$:

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

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 $T \rightarrow \infty$,

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

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

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

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

$$\text{plim}_{(T, N \rightarrow \infty)_{\text{seq}}} (\widehat{B} - B) = 0.$$

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 $T \rightarrow \infty$ to obtain:

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

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

$$\text{plim}_{(T, N \rightarrow \infty)_{\text{seq}}} (\hat{\eta}_t - \eta_t) = 0.$$

Finally,

$$\hat{\varepsilon}_{i,t} = f_{i,t}^h - y_t - \hat{k}_i - \hat{\eta}_t, \text{ and } f_{i,t}^h - y_t = k_i + \eta_t + \varepsilon_{i,t}.$$

Hence :

$$\hat{\varepsilon}_{i,t} - \varepsilon_{i,t} = (k_i - \hat{k}_i) + (\eta_t - \hat{\eta}_t).$$

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

$$\text{plim}_{(T, N \rightarrow \infty)_{\text{seq}}} (\hat{\varepsilon}_{i,t} - \varepsilon_{i,t}) = 0.$$

■

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 $\frac{1}{N} \sum_{i=1}^N f_{i,t}^h - \hat{B}$ obeys $\text{plim}_{(T, N \rightarrow \infty)_{\text{seq}}} \left(\frac{1}{N} \sum_{i=1}^N f_{i,t}^h - \hat{B} \right) = y_t + \eta_t = \mathbb{E}_{t-h}(y_t)$ and has a mean-squared error as follows:*

$\mathbb{E} \left[\text{plim}_{(T, N \rightarrow \infty)_{\text{seq}}} \left(\frac{1}{N} \sum_{i=1}^N f_{i,t}^h - \hat{B} \right) - y_t \right]^2 = \sigma_\eta^2$. Therefore it is an optimal forecasting device.

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

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

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

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

from (2) and (3), which is the optimal forecast. The MSE of the feasible bias-corrected average forecast is:

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

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

Here, combining forecasts using equal weights $1/N$ and a feasible bias-correction term is optimal, and we can approximate $\mathbb{E}_{t-h}(y_t)$ well enough¹¹. As before, any other forecast combination as in Corollary 4 will also be optimal. Again, in the limit, 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 $N \rightarrow \infty$, 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} . Hence, we need not limit the asymptotic path of N, T as was the case with optimal weights.

From a different angle, bias-correction can be viewed as a form of intercept correction as discussed in Palm and Zellner (1992) and Hendry and Mizon (2005), for example. From (16), we could retrieve \widehat{B} from an OLS

¹¹A key issue for a WLLNs to hold is that forecasts have enough diversity in the precise sense given in Assumption 4. In a recent paper, Clark and McCracken (2007) discuss how to compute optimal weights in combining nested models. If nested models are taken to mean *similar* models, in the sense that their individual forecast errors components $\varepsilon_{i,t}$ are highly correlated, then, Assumption 4 may be violated. Computing the minimal set of conditions under which a WLLNs holds for nested models is out of the scope of this paper.

regression of the form:

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

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 $\hat{\delta} = \hat{B}$, where \hat{B} is now cast in terms of the previous literature¹².

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:*

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

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

Proof. Under $H_0 : B = 0$, we have shown in Proposition 5 that \hat{B} is a $(T, N \rightarrow \infty)_{\text{seq}}$ consistent estimator for B . To compute the consistent estimator of the asymptotic variance of \bar{B} we follow Conley(1999), who matches spatial dependence to a metric of economic distance. Denote by $\text{MSE}_i(\cdot)$

¹²Recall that $\hat{B} = \frac{1}{N} \sum_{i=1}^N \hat{k}_i$, where only one \hat{k}_i is estimated separately using R observations. For weighted combinations – intercept and estimated weights $\hat{\omega}_i$, $i = 1, 2, \dots, N$ – joint estimation of $N + 1$ parameters is performed using these same R observations. This links the number of forecasts and the cross-sectional sample size for a given R , which does not happen for the feasible BCAF.

Our estimator is also less restrictive from an asymptotic point-of-view. For the weighted forecast combination, recall that the feasibility condition required that,

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

For the feasible BCAF, $c \in [0, \infty]$.

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 $i = 2, 3, \dots, N$. With this measure of spatial dependence at hand, we can construct a two-dimensional estimator of the asymptotic variance of \bar{B} and \hat{B} following Conley(1999, Sections 3 and 4). We label \bar{V} and \hat{V} the estimates of the asymptotic variances of \bar{B} and of \hat{B} , respectively.

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

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

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

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

■

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 of the target variable plus an additive bias term and an idiosyncratic error. Our DGP is a simple stationary $AR(1)$ process:

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

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 terms ($\varepsilon_{i,t}$). Because $\xi_t \sim \text{i.i.d.}N(0, 1)$, the optimal theoretical 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., $E = T_1 = 200$, so that $\hat{\alpha}_0 \simeq \alpha_0$ and $\hat{\alpha}_1 \simeq \alpha_1$. In practice, however, forecasters may have economic incentives to make biased forecasts, and there may be other sources of misspecification arising from misspecification errors. Therefore, we generate forecasts as:

$$\begin{aligned} f_{i,t} &= \hat{\alpha}_0 + \hat{\alpha}_1 y_{t-1} + k_i + \varepsilon_{i,t}, \\ &= (\hat{\alpha}_0 + k_i) + \hat{\alpha}_1 y_{t-1} + \varepsilon_{i,t} \text{ for } t = T_1 + 1, \dots, T, i = 1, \dots, N, \end{aligned} \quad (20)$$

where, $k_i = \beta k_{i-1} + u_i$, $u_i \sim \text{i.i.d. Uniform}(a, b)$, $0 < \beta < 1$, and $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{N,t})'$, $N \times 1$, is drawn from a multivariate Normal distribution with size $R + P = T - T_1$, whose mean vector equals zero and covariance matrix equals Σ . We introduce heterogeneity and spatial dependence in the distribution of $\varepsilon_{i,t}$. The diagonal elements of $\Sigma = (\sigma_{ij})$ 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 values of the σ_{ii} 's are randomly determined through an once-and-for-all draw from a uniform random variable of size N , that is, $\sigma_{ii} \sim \text{i.i.d. Uniform}(1, \sqrt{10})$ ¹³.

In equation (20), we built spatial dependence in the bias term k_i ¹⁴. The cross-sectional average of k_i is $\frac{a+b}{2(1-\beta)}$. We set the degree of spatial dependence in k_i by letting $\beta = 0.5$. For the support of u_i , we consider 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 it is $B = 0$ in (ii). Finally, notice that the specification of $\varepsilon_{i,t}$ satisfies Assumption 4 in Section 2 as we let $N \rightarrow \infty$.

Equation (20) is used to generate three panels of forecasts. They differ from each other in terms of the number of forecasters (N): $N = 10, 20, 40$.

¹³The covariance matrix Σ does not change over simulations.

¹⁴The additive bias k_i is explicit in (20). 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, in their paper, intercept correction is equivalent to bias correction, which would be the case here too. However, a structural break would violate weak stationarity and that is why it is not attempted here.

We assume that they all have the same training-sample and out-of-sample observations: $R = 50$, and $P = 50$, respectively. For each N , we conduct 50,000 simulations in the experiment, where the total number of time observations equals 300 in each panel ($E = T_1 = 200$, $R = 50$, and $P = 50$).

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. For these methods, our results include aspects of the whole distribution of their respective biases and MSEs.

For the BCAF, we use the training-sample observations to estimate $\widehat{k}_i = \frac{1}{R} \sum_{t=T_1+1}^{T_2} (y_t - f_{i,t})$ and $\widehat{B} = \frac{1}{N} \sum_{i=1}^N \widehat{k}_i$. Then, we compute the out-of-sample forecasts $\widehat{f}_t^{BCAF} = \frac{1}{N} \sum_{i=1}^N f_{i,t} - \widehat{B}$, $t = T_2 + 1, \dots, T$, and we employ the last P observations to compute $MSE_{BCAF} = \frac{1}{P} \sum_{t=T_2+1}^T (y_t - \widehat{f}_t^{BCAF})^2$.

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

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

where the restriction $\sum_{i=1}^N \omega_i = 1$ is imposed in estimation. The weighted forecast is $\widehat{f}_t^{weighted} = \widehat{\delta} + \widehat{\omega}_1 f_{1,t} + \widehat{\omega}_2 f_{2,t} + \dots + \widehat{\omega}_N f_{N,t}$, and the intercept δ plays the role of bias correction. We employ the last P observations to compute $MSE_{weighted} = \frac{1}{P} \sum_{t=T_2+1}^T (y_t - \widehat{f}_t^{weighted})^2$.

For the average forecast, there is no parameter to be estimated using training-sample observations. Out-of-sample forecasts are computed according to $f_t^{average} = \frac{1}{N} \sum_{i=1}^N f_{i,t}$, $t = T_2 + 1, \dots, T$, and its MSE is computed as $MSE_{average} = \frac{1}{P} \sum_{t=T_2+1}^T (y_t - f_t^{average})^2$.

Finally, for each approach, we also computed the out-of-sample mean biases. In small samples, the weighted forecast and the BCAF should have

out-of-sample mean biases close to zero, whereas the mean bias of the average forecast should be close to $B = \frac{a+b}{2(1-\beta)}$.

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, and (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 mean 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 a mean MSE at least 8.7% higher than that of the bias-corrected average forecast, 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 as 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 if N/R is close to unity from below. 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. The behavior of the weighted average forecast is identical to that in Table 1.

4 Empirical Application

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 , the Focus Survey 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, 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 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 $R = 26$ time observations to compute \hat{B} – the average bias – leaving $P = 18$ 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. First, we note that all the 18 individual forecasts perform worse than combinations, which is consistent with the discussion in Hendry and Clements (2002). The results in Table 3 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. In Table

4, out-of-sample forecast comparisons between the simple average and the bias-corrected average forecast show that the former has a MSE 18.2% bigger than that of the latter. We also computed the MSE of the weighted forecast. Since we have $N = 18$ and $R = 26$, $N/R = 0.69$. Hence, the weighted average cannot avoid the curse of dimensionality, yielding a MSE 390.2% bigger than that of the BCAF.

It is important to stress that, although the bias-corrected average forecast was conceived for a large N, T environment, the empirical results here show an encouraging performance even in a small N, T context. Also, the forecasting gains from bias correction are non-trivial.

5 Conclusions and Extensions

In this paper, we propose a novel approach to econometric forecast of stationary and ergodic series y_t within a panel-data framework, where the number of forecasts and the number of time periods increase without bounds. The basis of our method is a two-way decomposition of the forecasts error. As shown here, this is equivalent to forecasters trying to approximate the optimal forecast under quadratic loss – the conditional expectation $\mathbb{E}_{t-h}(y_t)$, which is modelled as the *common feature* of all individual forecasts. Standard tools from panel-data asymptotic theory are used to devise an optimal forecasting combination that delivers $\mathbb{E}_{t-h}(y_t)$. This optimal combination uses equal weights and an estimated 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 an estimated bias-correction term eliminates any possible detrimental effect arising from equal weighting. We label this optimal forecast as the (feasible) *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. 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. There may 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. In our Monte-Carlo simulations, when we consider $N = 40$, the simple average has a mean MSE 17.8% higher than that of the feasible bias-corrected average forecast, while the weighted combination has a mean MSE 431.3% larger. In the empirical exercise, for N as low as 18, the feasible bias-corrected average forecast leads to a sizable improvement in forecasting accuracy under MSE loss – from 18% to about 390% – respectively regarding the simple and the weighted average.

As a by-product of the use of panel-data asymptotic methods, with $N, T \rightarrow \infty$, we advanced the understanding of the forecast combination puzzle. The key issue is that simple averaging requires no estimation of weights, while optimal weights requires estimating N weights that grow unbounded in the asymptotic setup. We show that there is no forecast-combination puzzle under certain asymptotic paths for N and T , but not for all. Indeed, if $N \rightarrow \infty$ at a rate strictly smaller than that of T , and $\lim_{N, T \rightarrow \infty} N/T < \kappa_2$, then, the estimators of the weights are consistent, the weighted forecast with bias correction (intercept) is optimal, and there is no puzzle. Under different paths, it is impossible to obtain consistent estimators for the weights. For these paths, weights are not identified, their estimates are unfeasible, and the variance of the weighted forecast diverges to infinity. The Monte-carlo exercise illustrates a portion of such a path in finite samples. There, the variance of the weighted forecast increases with N when R is fixed. If N/R is close to unity from below, the variance component of the MSE of the weighted forecast will be large – larger, the closer N/R is to unity – and the simple average will be more accurate. This is the curse of dimensionality as an explanation to the puzzle.

References

- [1] Amemiya, T. (1971), “The estimation of the variances in the variance-components model”, *International Economic Review*, vol. 12, pp. 1-13.
- [2] Baltagi, Badi H., 1980, “On Seemingly Unrelated Regressions with Error Components,” *Econometrica*, Vol. 48(6), pp. 1547-1551.

- [3] Bai, J., (2005), "Panel Data Models with Interactive Fixed Effects," Working Paper: New York University.
- [4] Bai, J., and S. Ng, (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.
- [5] Bates, J.M. and Granger, C.W.J., 1969, "The Combination of Forecasts," *Operations Research Quarterly*, vol. 20, pp. 309-325.
- [6] Batchelor, R., 2007, "Bias in macroeconomic forecasts," *International Journal of Forecasting*, vol. 23, pp. 189-203.
- [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] Clark, T.E. and McCracken, M.W., 2007, "Combining Forecasts for Nested Models," Working Paper: Kansas City FED, forthcoming in the *Journal of Econometrics*.
- [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. Korajczyk (1986), "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373-394.
- [12] Davies, A. and Lahiri, K., 1995, "A new framework for analyzing survey forecasts using three-dimensional panel data," *Journal of Econometrics*, vol. 68(1), pp. 205-227
- [13] Elliott, G., C.W.J. Granger, and A. Timmermann, 2006, Editors, *Handbook of Economic Forecasting*, Amsterdam: North-Holland.
- [14] Elliott, G. and A. Timmermann (2005), "Optimal forecast combination weights under regime switching", *International Economic Review*, 46(4), 1081-1102.

- [15] Elliott, G. and A. Timmermann (2004), “Optimal forecast combinations under general loss functions and forecast error distributions”, *Journal of Econometrics* 122:47-79.
- [16] Engle, R.F. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50, pp. 987-1006.
- [17] Engle, R. F., Issler, J. V., 1995, “Estimating common sectoral cycles,” *Journal of Monetary Economics*, vol. 35, 83–113.
- [18] Engle, R.F. and Kozicki, S. (1993). “Testing for Common Features”, *Journal of Business and Economic Statistics*, 11(4): 369-80.
- [19] Forni, M., Hallin, 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.
- [20] Forni M., Hallin M., Lippi M. and Reichlin L., 2003 “The Generalized Dynamic Factor Model one-sided estimation and forecasting,” *Journal of the American Statistical Association*, forthcoming.
- [21] 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.
- [22] Granger, C.W.J., 1989, “Combining Forecasts-Twenty Years Later,” *Journal of Forecasting*, vol. 8, 3, pp. 167-173.
- [23] Granger, C.W.J., and R. Ramanathan (1984), “Improved methods of combining forecasting”, *Journal of Forecasting* 3:197–204.
- [24] Hendry, D.F. and M.P. Clements (2002), “Pooling of forecasts”, *Econometrics Journal*, 5:1-26.
- [25] Hendry, D.F. and Mizon, G.E. (2005): “Forecasting in the Presence of Structural Breaks and Policy Regime Shifts”, in “*Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*,” D.W.K. Andrews and J.H. Stock (eds.), Cambridge University Press.

- [26] 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.
- [27] 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.
- [28] Kang, H. (1986), "Unstable Weights in the Combination of Forecasts," *Management Science* 32, 683-95.
- [29] Laster, David, Paul Bennett and In Sun Geoum, 1999, "Rational Bias In Macroeconomic Forecasts," *The Quarterly Journal of Economics*, vol. 114, issue 1, pp. 293-318
- [30] Levin, A. and Lin, C.F. (1993), "Unit root tests in panel data: asymptotic and finite-sample properties," Discussion paper, University of California, San Diego.
- [31] 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.
- [32] Patton, Andrew J. and Allan Timmermann, 2006, "Testing Forecast Optimality under Unknown Loss," forthcoming in the *Journal of the American Statistical Association*.
- [33] Pesaran, M.H., (2005), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." Working Paper: Cambridge University, forthcoming in *Econometrica*.
- [34] Phillips, P.C.B. and H.R. Moon, 1999, "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, vol. 67 (5), pp. 1057–1111.
- [35] Quah, D. (1994), "Exploiting cross-section variation for unit root inference in dynamic data," *Economics Letters*, 44, pp. 9-19.
- [36] Stock, J. and Watson, M., "Forecasting Inflation", *Journal of Monetary Economics*, 1999, Vol. 44, no. 2.

- [37] Stock, J. and Watson, M., “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business and Economic Statistics*, April 2002a, Vol. 20 No. 2, 147-162.
- [38] Stock, J. and Watson, M., “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 2002b.
- [39] 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.
- [40] 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.
- [41] Vahid, F. and Engle, R. F., 1993, “Common trends and common cycles,” *Journal of Applied Econometrics*, vol. 8, 341–360.
- [42] Vahid, F., Engle, R. F., 1997, “Codependent cycles,” *Journal of Econometrics*, vol. 80, 199–221.
- [43] 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.
- [44] Wallace, H. D., Hussain, A., 1969, “The use of error components model in combining cross-section and time-series data,” *Econometrica*, 37, 55–72.
- [45] West, K., 1996, “Asymptotic Inference about Predictive Ability,” *Econometrica*, 64, (5), pp. 1067-84.

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
τ -th= 0.01 quantile	-0.590	-0.091	-0.643	0.911	0.989	1.080
τ -th= 0.25 quantile	-0.169	0.252	-0.186	1.337	1.448	1.608
τ -th= 0.50 quantile	-0.001	0.391	-0.001	1.540	1.672	1.872
τ -th= 0.75 quantile	0.167	0.530	0.184	1.763	1.918	2.179
τ -th= 0.99 quantile	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
τ -th= 0.01 quantile	-0.532	-0.001	-0.690	0.754	0.853	1.117
τ -th= 0.25 quantile	-0.153	0.316	-0.195	1.098	1.247	1.723
τ -th= 0.50 quantile	-0.001	0.440	-0.004	1.266	1.444	2.053
τ -th= 0.75 quantile	0.151	0.565	0.192	1.452	1.659	2.448
τ -th= 0.99 quantile	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
τ -th= 0.01 quantile	-0.515	0.050	-1.209	0.673	0.786	2.165
τ -th= 0.25 quantile	-0.145	0.346	-0.315	0.980	1.150	4.021
τ -th= 0.50 quantile	0.000	0.465	0.002	1.130	1.331	5.337
τ -th= 0.75 quantile	0.141	0.583	0.312	1.295	1.529	7.243
τ -th= 0.99 quantile	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
τ -th= 0.01 quantile	-0.590	-0.511	-0.643	0.911	0.905	1.080
τ -th= 0.25 quantile	-0.169	-0.149	-0.186	1.337	1.326	1.608
τ -th= 0.50 quantile	-0.001	0.000	-0.001	1.540	1.526	1.872
τ -th= 0.75 quantile	0.167	0.147	0.184	1.763	1.745	2.179
τ -th= 0.99 quantile	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
τ -th= 0.01 quantile	-0.532	-0.462	-0.690	0.754	0.746	1.117
τ -th= 0.25 quantile	-0.153	-0.130	-0.195	1.098	1.089	1.723
τ -th= 0.50 quantile	-0.001	0.000	-0.004	1.266	1.254	2.053
τ -th= 0.75 quantile	0.151	0.130	0.192	1.452	1.435	2.448
τ -th= 0.99 quantile	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
τ -th= 0.01 quantile	-0.515	-0.426	-1.209	0.673	0.667	2.165
τ -th= 0.25 quantile	-0.145	-0.123	-0.315	0.980	0.971	4.021
τ -th= 0.50 quantile	-0.002	0.000	0.002	1.130	1.116	5.337
τ -th= 0.75 quantile	0.141	0.121	0.312	1.295	1.278	7.243
τ -th= 0.99 quantile	0.509	0.424	1.209	1.771	1.733	17.669

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

Horizon (h)	Avg. Bias \hat{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 Average 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.