Time-varying bias-corrected average forecast

Gilberto Boaretto
UERJ, PUC-Rio & Ibmec/RJ, Brazil
gilbertoboaretto@hotmail.com

Marcelo Medeiros
UIUC, US
marcelom@illinois.edu

This draft: March 20, 2026

Abstract

We propose a bias correction for the average of a set of individual inflation expectations considering the possibility that intercept and slope biases may vary over time. We proceed in two ways. Firstly, we consider estimations based on rolling windows. Secondly, we employ a state-space model to obtain time-varying intercept and slope biases using the recursiveness of the Kalman filter. The latter approach has the advantage of circumventing the choice of the rolling window size. We also proceed with estimations based on expanding windows, a procedure that is close to what has been done in the literature. We achieve good forecast performance for models based on small rolling windows for shorter and intermediate forecast horizons. In turn, a state-space model that includes corrections for intercept and slope biases varying over time tends to perform slightly worse than procedures based on rolling windows.

Keywords: inflation forecasting; combination of forecasts; bias correction; time-varying parameters; Kalman filter.

JEL Codes: C32, C38, C51, C52, C53, E37.

XXVI Brazilian Finance Meeting (EBFin)

Fortaleza, Brazil – July 22-24, 2026

Track: Financial Econometrics

1 Introduction

It is unlikely that some individual forecasting model will consistently outperform all others over time and in different economic conditions since models likely present inaccuracies in their specifications (Elliott and Timmermann, 2016). The aim of combining forecasts is to reduce uncertainty and increase the accuracy of the last forecast. Bates and Granger (1969) show that the combination of *unbiased* forecasts could yield lower mean-squared error than either of the original forecasts. More specifically, they demonstrate that two unbiased forecasts can be optimally combined to obtain a variance no greater than the smaller of the two individual variances. Reid (1969) extends this approach by proposing the combination of more than two unbiased forecasts, and Newbold and Granger (1974) achieve good results by applying this procedure. Granger and Ramanathan (1984) add an intercept in a linear model that combines *potentially biased* individual forecasts. However, there is a challenge when we deal with a large number of individual forecasts and few time observations. The estimation uncertainty could compromise the results, or we might not even be able to estimate the model if we have more individual forecasts than observations.

A solution adopted in the literature to combine multiple potentially biased forecasts avoiding over-parameterization has been to use the average forecast directly. After decomposing individual forecast errors into time-fixed forecast bias, time-varying aggregate uncertainty of the forecasts, and idiosyncratic terms, Palm and Zellner (1992) average equations over cross-sectional dimension and obtains a combination that corrects for additive (intercept) bias. Using panel-data sequential asymptotics, Issler and Lima (2009) show that a bias-correction average forecast (BCAF) obtained via average forecast error is equivalent to the conditional expectation and has an optimal limiting mean-squared error. However, if there is not only additive but also multiplicative (slope) bias, the correction only for intercept bias (BCAF) is no longer optimal. Capistrán and Timmermann (2009) considers both additive and multiplicative bias for bias-adjusting the equal-weighted forecast. Gaglianone and Issler (2023) propose an extended bias-corrected average forecast (EBCAF), which also considers the correction for both types of biases. Additionally, they highlight the implications of the existence of public and private information for the combination of forecasts.

We propose time-varying (extended) bias-corrected average forecast models (TV-BCAF and TV-EBCAF), which account for intercept and slope biases varying over time. Initially, we consider some time variation by means of OLS estimation based on rolling windows. We apply the procedure to different sets of individual forecasts for Brazilian inflation, including the median of available inflation expectations from the Focus survey (the Focus consensus) and forecasts generated by models discussed in **?**. We find that rolling-window-based models perform well, particularly for short windows when forecasting inflation at shorter and intermediate horizons (one to six months ahead). We then propose a state-space model that uses the recursiveness of the Kalman filter to obtain time-varying intercept and slope biases. This approach avoids the need for a discretionary choice of the rolling window size. Overall, models based on the Kalman filter including both types of biases perform slightly worse than some rolling window-based procedures. However, we encourage exploring variations in the specification of the state-space models and finding alternatives for reducing the variance of the estimated time-varying biases.

Outline. This paper comprises four additional sections following this Introduction. Section 2 outlines some extant procedures in the literature for correcting biases and presents their results. Section 3 describes models considering time-varying bias correction for both intercept and slope. In Section 4, we present and analyze the forecast performances of models that incorporate time-varying bias, while comparing them to traditional combination methods. Section 5 brings the final considerations of this paper. Appendix A provides a brief description of the Kalman filter while Appendix B contains figures displaying the temporal evolution of intercept bias estimated via different approaches.

2 Methodologies for combining forecasts

For a given period T and a forecast horizon of h months ahead, consider a set of N inflation forecasts designated by $\{\widehat{\pi}_{i,T+h\,|\,T} : i = 1, \ldots, N\}$. Each individual forecast is generated by a model or comes from a survey of experts, for example. For simplicity, we will treat all original forecasts as coming from an unknown model. For a survey of forecasts, this is natural. For projections originating from estimated models, this is a strong simplification. However, it does away with the difficulties inherent in considering a combination of *models* rather than a combination of *forecasts* whose generating process is unknown. Furthermore, the practitioner wants to know whether a combination of forecasts generates empirically satisfactory results.

2.1 Average forecast

The simplest way to combine forecasts is to compute the average of the available forecasts. Thus, we compute

$$\widehat{\pi}^{\text{av}}_{T+h\,|\,T} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\pi}_{i,T+h\,|\,T}. \tag{1}$$

The main advantage of this approach is that it does not require the estimation of weights for available forecasts, which would require training-sample observations. In this setup, the big challenge would be to obtain stable weights over time since we have a small sample of individual forecasts. On the other hand, in the average forecast, the weights assigned to individual forecasts are equal to $\frac{1}{N}$. However, one of the main drawbacks of the average forecast is that assigning the same weight to inaccurate individual forecasts can lead to a sub-optimal combined forecast. That might be true if the set of individual forecasts includes biased forecasts.

2.2 A bias-corrected average forecast (BCAF)

We can make a modification to achieve an unbiased combined forecast, which is particularly useful when we potentially consider biased forecasts within the pool of individual forecasts. As in Issler and Lima (2009), consider $\mathbb{E}_{t-h}(\pi_t) = \mathbb{E}(\pi_t \mid \mathcal{F}_{t-h})$, the expectation of π_t conditional to a information set available at $t-h$, $\mathcal{F}_{t-h}$, is an optimal device. Then, the econometrician's aim is to approximate this function.

Two-way decomposition or error-component decomposition. Let us consider that an individual forecast $\widehat{\pi}_{i,t\,|\,t-h}$ aims to approximate $\mathbb{E}_{t-h}(\pi_t)$. Thus, we can define the approximation error as

$$\mathbb{E}_{t-h}(\pi_t) - \widehat{\pi}_{i,t\,|\,t-h} = \delta_i^h + \varepsilon_{i,t}^h, \qquad i = 1, \dots, N, \tag{2}$$

where δ_i^h is the individual model time-invariant bias considering the forecast horizon h, and $\varepsilon_{i,t}^h$ is the individual model error term in approximating $\mathbb{E}_{t-h}(\pi_t)$ with $\mathbb{E}(\tilde{\varepsilon}_{i,t}^h) = 0$ for all i, t, and h. In turn, consider the error for conditional expectation $\mathbb{E}_{t-h}(\pi_t)$ given by

$$\pi_t - \mathbb{E}_{t-h}(\pi_t) = \theta_t^h \tag{3}$$

where θ_t^h is an unpredictable time-component with $\mathbb{E}_{t-h}(\theta_t^h) = 0$ for all t and h. Finally, combining (2) and (3), we obtain the forecast error

$$\pi_t - \widehat{\pi}_{i,t\,|\,t-h} = \delta_i^h + \theta_t^h + \varepsilon_{i,t}^h. \tag{4}$$

The term δ_i^h captures a fixed long-term effect on the forecast generated by a model or survey respondent. The term θ_t^h captures time effects arising from the lack of future information between $t-h$ and t, which equally affects all models or respondents. Finally, the term $\varepsilon_{i,t}^h$ captures idiosyncratic errors that affect individuals differently over time (Issler and Lima, 2009).

Issler and Lima's bias-corrected average forecast. Consider the following assumptions:

(i) δ_i^h, θ_t^h, and $\varepsilon_{i,t}^h$ are independent of each other for all i and t;

(ii) δ_i^h is an identically distributed random variable in the cross-sectional dimensional i with mean δ^h and variance σ_δ^2;

(iii) θ_t^h is a stationary and ergodic MA process of order at most $h-1$ with zero mean and finite variance;

(iv) limited degree of cross-sectional dependence of erros, $\varepsilon_{i,t}^h$.

Under these assumptions, Issler and Lima (2009) show that a bias-corrected average forecast (BCAF) given by

$$\widehat{\pi}^{\text{BCAF}}_{T+h\,|\,T} = \widehat{\delta}^h + \widehat{\pi}^{\text{av}}_{T+h\,|\,T}, \tag{5}$$

where

$$\widehat{\delta}^h = \frac{1}{N}\sum_{i=1}^{N}\widehat{\delta}^h_i, \qquad \text{with} \qquad \widehat{\delta}^h_i = \frac{1}{T}\sum_{t=1}^{T}\pi_t - \frac{1}{T}\sum_{t=1}^{T}\widehat{\pi}_{i,t-h\,|\,t}, \quad i = 1,\ldots,N.$$

is an optimal forecasting device. Notice that $\widehat{\pi}^{\text{av}}_{T+h\,|\,T}$ is the average forecast defined in (1).

For the sake of convenience, we define the forecast error in the usual way, i.e., as $y - \widehat{y}$, where y is the actual variable and $\widehat{y}$ is a forecast for y. Note that Issler and Lima (2009) proceed in an inverse way, as is usual in the literature on error-component decomposition (e.g. Palm and Zellner, 1992). For the model just presented, this difference is not relevant. The relevant difference, which appears in the presentation of the results in Subsection 2.3, is that we employ rolling windows of different sizes (including very short ones) to obtain the intercept bias. The procedure suggested by Issler and Lima (2009) assumes that both N and T diverge, which is not compatible with limiting ourselves to using only part of the sample to obtain the bias. However, we also obtain the intercept bias using data in an extended window, that is, considering all available information, which is closer to Issler and Lima's procedure.

Extended error-component decomposition. Now, consider an extended error-component decomposition in which besides intercept bias, we consider the possibility that there is a slope bias as well. Therefore, the model is able to capture both additive and multiplicative biases. In this setup, a decomposition can be written as

$$\mathbb{E}_{t-h}(\pi_t) - \beta^h\,\widehat{\pi}_{i,t\,|\,t-h} = \delta^h_i + \varepsilon^h_{i,t}, \qquad i = 1,\ldots,N \tag{6}$$

$$\pi_t - \mathbb{E}_{t-h}(\pi_t) = \theta^h_t. \tag{7}$$

By combining the Equations (6) and (7), we obtain a forecast error given by

$$\pi_t - \beta^h\,\widehat{\pi}_{i,t\,|\,t-h} = \delta^h_i + \theta^h_t + \varepsilon^h_{i,t}, \qquad i = 1,\ldots,N. \tag{8}$$

Averaging (8) over the cross-sectional dimension i and solving for π_t yields

$$\pi_t = \delta^h + \beta^h\,\widehat{\pi}^{\text{av}}_{t\,|\,t-h} + u^h_t, \tag{9}$$

where $u^h_t = \theta^h_t + \varepsilon^h_t$.

Model (9) is exactly a model estimated in Capistrán and Timmermann (2009), basically a bias adjustment of the equal-weight forecast. We can also interpret it as an aggregated version of the Mincer and Zarnowitz's (1969) equation used as a first step to test the rationality of a forecast. In this case, rationality is corroborated under $\delta^h = 0$ and $\beta^h = 1$, for all h. Thus, we can proceed with a test for the hypothesis of the rationality of expectations. Several papers point to violating the rationality hypothesis. Then, a procedure that adjusts both intercept and slope bias can be very useful for obtaining an unbiased final forecast. In this line, for each horizon forecast h, an extended bias-corrected average forecast (EBCAF) is given by

$$\widehat{\pi}^{\text{EBCAF}}_{T+h\,|\,T} = \widehat{\delta}^h + \widehat{\beta}^h \, \widehat{\pi}^{\text{av}}_{T+h\,|\,T}, \tag{10}$$

where $\widehat{\delta}^h$ and $\widehat{\beta}^h$ are OLS estimates.

The approach to intercept and slope bias-correction presented here differs from that suggested by Gaglianone and Issler (2023). They also assume that forecasts obey a factor model with an affine structure and derive a model in which the slope coefficient appears by multiplying the conditional expectation of π_t rather than the average forecast. Suggested by Issler and Lima (2009), this approach is used in Gaglianone, Issler, and Matos (2017) that, as recommended by Gaglianone and Issler (2023), stacking the models along the different forecast horizons to proceed with the estimation using GMM, which can lead to efficiency gains. Another advance of Gaglianone and Issler (2023) in relation to Issler and Lima (2009) is to show that the procedure is consistent with T diverging and N being kept fixed. Finally, due to the possibility of the existence of private information, Gaglianone and Issler show that the average forecast will be correlated with the error term, so that the aggregate Mincer-Zarnowitz regression should be estimated using instrumental variables. However, since in this essay we use the median of the available Focus and forecasts based on public information, this limits the possibility of endogeneity through the channel raised by the authors. Along these lines, the main theoretical problem with our approach remains the fact that we know the generating process of model-based forecasts.

2.3 Results of a pseudo-out-of-sample forecast exercise

Setup. We set January 2015 as the starting point for generating combined forecasts to ensure we have at least 25 time periods available when we use a bias-corrected average forecast (BCAF) based on extending windows since individual forecasts began in January 2012[1] (see **?**). We explore BCAF models implemented with rolling windows of different lengths: 3, 6, 7, 8, 9, 10, 12, and 24 months. However, to avoid redundancy, besides the average forecast, and BCAF and EBCAF models based on expanding windows, we report results for BCAF based on rolling windows of 3, 6, 8, 9, and 12 months. We choose the root mean squared error (RMSE) as the metric to evaluate the accuracy of forecasting results. All out-of-sample

[1] Initially, there are a total of 25 periods available for $h = 12$ since the 12-month-ahead forecast for Jan/2015 is calculated in Feb/2014. On the other hand, note that there are 36 available periods for $h = 0$.

RMSE are normalized with respect to the median of the available Focus survey's inflation expectations. We also report the RMSE ratio of the median of the *ex-post* Focus survey's inflation expectations.

Forecast performance. Table 1 presents results for the period from January 2015 to June 2022. Forecast performance varies depending on the forecast horizon and the rolling window length. For nowcasting ($h = 0$), no forecast combination outperforms the median of ex-post Focus' inflation expectations (Panel A). For longer horizons ($h \geqslant 6$), the average forecast (Panel B) exhibits the best performances. For intermediate horizons ($1 \leqslant h \leqslant 7$), the BCAF models based on rolling windows of 6, 8, 9, and 12 months (Panels D to G) register good predictive performances. Notably, for this subset of horizons and rolling window sizes, the predictive accuracy for shorter horizons improves as the rolling window size increases.

However, choosing the rolling window size can be challenging for the econometrician. In addition, it is important to note that, despite being numerically superior for some horizons, most of the bias-correction-based combinations results are not statistically significant compared to the predictive performance of the median of available Focus' expectations, according to Diebold-Mariano tests. These forecasts would also not be statistically superior to the average forecast without bias correction. Despite this, the results highlight the importance of considering the possibility of bias in the average forecast for short and intermediate horizons, as well as the possibility that this bias varies over time. The lack of statistical significance may be associated with estimation uncertainty, which is enlarged when considering fewer observations, which suggests a trade-off in defining the rolling window size: a smaller window can better capture variations in time but can increase the estimation instability.

Regarding the set of individual forecasts employed, the set comprising forecasts specifically generated for aggregate inflation, which includes the median of Focus' expectations (displayed in the initial rows of Panels B to I in Table 1), yields the most favorable forecast outcomes compared to alternative sets of individual forecasts. In certain instances, the use of indirect forecasts formed through the aggregation of forecasts for BCB disaggregates (namely, administered, tradable, and non-tradable items) produces combined forecasts exhibiting predictive performance that closely approximates those obtained when utilizing the set of direct forecasts to compute the combinations.

Table 1: Out-of-sample RMSE ratio for average forecast and BCAF and EBCAF models with respect to the available Focus: Jan/2015 to Jun/2022

Method/Forecasts	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$	$h=11$
A. Survey												
Focus (available)	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*1.000*	*1.000*
Focus (*ex-post*)	**0.931*****	*0.972****	0.993***	1.001	1.000	1.000	1.000	0.999	0.999*	1.000	*1.001*	*1.001*
B. Average forecast												
Aggregates & Focus	1.215	*0.989*	0.953**	0.971*	0.982	0.970	*0.939***	**0.928*****	**0.930****	**0.935****	**0.968***	**0.986**
BCB & Focus	*1.208*	1.004	0.967	0.989	0.994	0.977	*0.945***	*0.951**	*0.948***	**0.950****	**0.973**	**0.985**
Groups & Focus	1.341	1.049	0.971	0.991	0.984	0.968	0.968	*0.950**	*0.953**	*0.976*	*0.975*	*0.992*
Subgroups & Focus	1.514	1.110	1.026	1.031	1.019	0.998	0.973	*0.960*	*0.977*	*0.996*	1.016	1.026
All & Focus	1.317	1.033	0.974	0.991	0.990	0.972	0.949**	**0.940****	**0.945***	*0.957**	*0.977*	*0.993*
C. Bias-corrected average forecast (BCAF) with rolling windows of 3 months												
Aggregates & Focus	1.340	1.116	1.087	1.096	1.083	1.041	0.982	0.972	0.995	1.047	1.122	1.183
BCB & Focus	1.378	1.132	1.077	1.100	1.067	1.028	0.967	0.967	0.993	1.045	1.118	1.146
Groups & Focus	1.499	1.198	1.106	1.138	1.093	1.049	1.012	0.977	0.984	1.068	1.130	1.163
Subgroups & Focus	1.627	1.240	1.152	1.142	1.097	1.055	1.010	0.979	1.008	1.079	1.161	1.206
All & Focus	1.463	1.175	1.107	1.121	1.084	1.041	0.989	0.972	0.994	1.058	1.130	1.173
D. Bias-corrected average forecast (BCAF) with rolling windows of 6 months												
Aggregates & Focus	1.261	1.029	0.979	0.969	*0.957*	*0.943*	**0.932**	*0.961*	0.994	1.040	1.096	1.135
BCB & Focus	1.283	1.056	0.974	0.977	*0.953*	**0.936**	**0.931**	0.977	1.009	1.049	1.106	1.116
Groups & Focus	1.407	1.108	1.001	1.006	0.968	**0.938**	*0.944*	0.969	1.008	1.091	1.121	1.136
Subgroups & Focus	1.562	1.157	1.045	1.021	0.983	0.957	0.951	0.974	1.035	1.103	1.157	1.183
All & Focus	1.380	1.091	1.000	0.992	0.962	*0.939*	*0.935*	0.968	1.010	1.069	1.118	1.142
E. Bias-corrected average forecast (BCAF) with rolling windows of 8 months												
Aggregates & Focus	1.235	*0.980*	*0.914**	**0.924**	**0.942**	*0.950*	*0.944*	*0.958*	0.991	1.035	1.089	1.128
BCB & Focus	1.266	1.003	*0.913**	*0.935*	**0.946**	0.955	0.952	0.983	1.012	1.050	1.100	1.114
Groups & Focus	1.380	1.053	0.937	0.958	*0.947*	*0.942*	0.961	0.975	1.016	1.094	1.117	1.139
Subgroups & Focus	1.531	1.098	0.982	0.975	0.965	0.965	0.963	0.976	1.043	1.108	1.157	1.186
All & Focus	1.354	1.035	*0.935*	*0.946*	*0.946*	*0.949*	0.951	0.971	1.014	1.070	1.114	1.141
F. Bias-corrected average forecast (BCAF) with rolling windows of 9 months												
Aggregates & Focus	1.209	**0.958**	**0.904***	**0.924**	*0.955*	0.958	*0.945*	0.964	0.993	1.039	1.093	1.127
BCB & Focus	1.239	*0.979*	**0.907***	*0.936*	0.961	0.966	0.955	0.991	1.015	1.052	1.103	1.116
Groups & Focus	1.355	1.030	*0.925*	*0.955*	*0.954*	*0.949*	0.961	0.983	1.019	1.097	1.120	1.140
Subgroups & Focus	1.500	1.074	0.972	0.973	0.973	0.972	0.963	0.985	1.047	1.113	1.163	1.185
All & Focus	1.327	1.011	*0.925*	*0.945*	0.957	0.958	0.951	0.978	1.016	1.074	1.118	1.141
G. Bias-corrected average forecast (BCAF) with rolling windows of 12 months												
Aggregates & Focus	*1.194*	*0.976*	*0.933*	*0.957*	0.983	0.979	0.968	0.988	1.019	1.058	1.109	1.144
BCB & Focus	1.215	0.995	0.935	0.967	0.991	0.992	0.981	1.016	1.040	1.071	1.122	1.137
Groups & Focus	1.331	1.043	0.953	0.976	0.973	0.967	0.986	1.009	1.044	1.113	1.137	1.158
Subgroups & Focus	1.468	1.084	0.998	0.996	0.994	0.993	0.991	1.012	1.074	1.133	1.179	1.196
All & Focus	1.303	1.025	0.954	0.972	0.982	0.980	0.977	1.004	1.042	1.092	1.135	1.158
H. Bias-corrected average forecast (BCAF) with expanding windows												
Aggregates & Focus	1.217	*0.987*	0.954*	0.972	0.986	0.976	0.953*	*0.948**	*0.957*	*0.969*	*1.005*	*1.025*
BCB & Focus	1.216	1.008	0.974	1.001	1.010	0.996	0.971	0.982	*0.981*	*0.986*	*1.012*	*1.024*
Groups & Focus	1.350	1.057	0.979	1.005	1.002	0.990	0.995	0.982	0.992	1.020	1.020	1.034
Subgroups & Focus	1.520	1.112	1.033	1.045	1.041	1.024	1.003	0.996	1.018	1.045	1.070	1.079
All & Focus	1.324	1.038	0.981	1.003	1.007	0.993	0.975	0.972	*0.982*	1.000	1.023	1.038
I. Extended bias-corrected average forecast (EBCAF) with expanding windows												
Aggregates & Focus	*1.074*	**0.967**	0.967	0.987	1.013	1.009	0.996	0.998	1.016	1.028	1.039	1.090
BCB & Focus	*1.139*	0.990	0.986	1.013	1.026	1.007	0.984	0.998	0.997	*0.996*	1.017	1.046
Groups & Focus	*1.208*	1.046	0.987	1.019	1.019	1.005	1.009	1.005	1.009	1.028	1.025	1.034
Subgroups & Focus	1.302	1.120	1.049	1.045	1.043	1.030	1.002	0.996	1.011	1.031	1.069	1.056
All & Focus	*1.153*	1.023	0.995	1.017	1.027	1.013	0.998	1.004	1.011	1.026	1.046	1.069

Notes: ***, **, and * indicate that for a specific forecast horizon, a forecast combination "comb" performed statistically better than the median of the available Focus at 1, 5, and 10% significance levels in a one-tailed Diebold-Mariano test with $\mathbb{H}_0 : \text{MSE}(\hat{\pi}^{\text{comb}}_{t+h\,|\,t}) = \text{MSE}(\pi^{\text{Focus}}_{t+h\,|\,t})$ *versus* $\mathbb{H}_1 : \text{MSE}(\hat{\pi}^{\text{comb}}_{t+h\,|\,t}) < \text{MSE}(\pi^{\text{Focus}}_{t+h\,|\,t})$. The two values highlighted in bold blue indicate the best and second-best methods for each horizon in terms of out-of-sample RMSE, while the six values in blue italics indicate the third- to eighth-best methods. All sets of individual forecasts included the median of the Focus survey's inflation expectations.

3 Time-varying bias correction for the average forecast

Analyzing the Brazilian Focus survey of forecasts for inflation, Carvalho and Minella (2012) note that there is empirical evidence for the existence of common forecast errors prevailing over idiosyncratic components among respondents. Moreover, they highlight the influence exerted by top-performing forecasters on other respondents, indicating a contamination phenomenon known as the epidemiology of the survey forecasts. Beyond that, the results of the preceding section reveal performance disparities in the bias-corrected average forecast when considering different sizes of rolling windows for bias estimation, as well as variations across forecast horizons. This prompts the hypothesis that biases vary over time, potentially possessing a certain degree of predictability. To verify this proposition, we introduce time-varying terms into the decomposition of individual forecasts. For the sake of simplicity and in line with Carvalho and Minella (2012), we assume that this term, which engenders temporal variations in biases, is common to all respondents and models. In practice, since we focus on the average forecast, questioning the adequacy of this hypothesis becomes a secondary concern.

3.1 Including a time-varying intercept bias

Error-component decomposition with time-varying intercept bias. Consider a new decomposition given by

$$\mathbb{E}_{t-h}(\pi_t) - \widehat{\pi}_{i,t\,|\,t-h} = \delta_i^h + \mu_t^h + \varepsilon_{i,t}^h, \qquad i = 1, \ldots, N \tag{11}$$

$$\mu_t^h = \mu_{t-h}^h + \nu_t^h \tag{12}$$

$$\pi_t - \mathbb{E}_{t-h}(\pi_t) = \theta_t^h \tag{13}$$

where we add a time-varying term μ_t common to all respondents or models in the individual forecast decomposition (Equation 11). Equation (12) indicates that this common term follows a random walk process – a common assumption in the literature on time-varying parameters –, and error term ν_t^h independent and identically distributed with zero mean. Lastly, notice that the decomposition of the conditional expectation using information available up to $t-h$ (Equation 13) is identical to the former (Equation 3). By combining the Equations (11) and (13), we obtain a forecast error given by

$$\pi_t - \widehat{\pi}_{i,t\,|\,t-h} = \delta_i^h + \mu_t^h + \theta_t^h + \varepsilon_{i,t}^h, \qquad i = 1, \ldots, N. \tag{14}$$

State-space representation. Averaging Equation (14) over i, we obtain a forecast error given by

$$\pi_t - \widehat{\pi}^{\text{av}}_{t\,|\,t-h} = \delta^h + \mu_t^h + \theta_t^h + \varepsilon_t^h, \tag{15}$$

where $\widehat{\pi}^{\text{av}}_{t\,|\,t-h} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\pi}_{i,t\,|\,t-h}$, such as defined in Equation (1), $\delta^h = \frac{1}{N}\sum_{i=1}^{N}\delta_i^h$, and $\varepsilon_t^h = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}^h$. By defining $\alpha_t^h = \delta^h + \mu_t^h$ and $u_t^h = \theta_t^h + \varepsilon_t^h$, and isolating π_t, we can rewrite Equations (15) and (12) in a state-space representation given by

$$\pi_t = \alpha_t^h + \widehat{\pi}^{\text{av}}_{t\,|\,t-h} + u_t^h \tag{16}$$

$$\alpha_t^h = \alpha_{t-h}^h + \nu_t^h \tag{17}$$

Notice that α_t^h is a time-varying average intercept bias, and u_t^h and ν_t^h are error terms.

For identification purposes, we complete the state-space representation assuming the following distribution for disturbances u_t^h and ν_t^h:

$$\begin{pmatrix} u_t^h \\ \nu_t^h \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0},\ \text{diag}\left(\sigma_u^2, \sigma_\nu^2\right)\right), \tag{18}$$

where we omitted the indication of the forecast horizon h in the variances.

Time-varying BCAF (TV-BCAF). Considering a sample with T temporal observations, we can estimate the system composed by the Equations (16), (17), and (18) by maximum likelihood using the Kalman filter recursion. A time-varying bias-corrected average forecast is given by

$$\widehat{\pi}^{\text{TV-BCAF}}_{T+h\,|\,T} = \widehat{\alpha}^h_{T+h\,|\,T} + \widehat{\pi}^{\text{av}}_{T+h\,|\,T},$$

where $\widehat{\alpha}^h_{T+h\,|\,T}$ is a *predict value* for state variable α^h_{T+h}, recovered using the Kalman filter.

3.2 Adding a time-varying slope bias

Extended error-component decomposition with time-varying intercept and slope bias. Lastly, consider a full error-component decomposition in which besides time-varying intercept bias, we consider the possibility that there is a time-varying slope bias as well. Therefore, the model is able to capture both additive and multiplicative bias. In this setup, the decomposition system is

$$\mathbb{E}_{t-h}(\pi_t) - \beta_t^h\,\widehat{\pi}_{i,t\,|\,t-h} = \delta_i^h + \mu_t^h + \varepsilon_{i,t}^h, \qquad i = 1, \ldots, N \tag{19}$$

$$\mu_t^h = \mu_{t-h}^h + \nu_t \tag{20}$$

$$\beta_t^h = \beta_{t-h}^h + \eta_t \tag{21}$$

$$\pi_t - \mathbb{E}_{t-h}(\pi_t) = \theta_t^h \tag{22}$$

where ν_t and η_t are independent error terms. Notice that we already assume, by parsimony, that both time-varying intercept and slope bias follow random walk processes.

By combining the Equations (19) and (22), we obtain a forecast error

$$\pi_t - \beta_t^h \, \widehat{\pi}_{i,t\,|\,t-h} = \delta_i^h + \mu_t^h + \theta_t^h + \varepsilon_{i,t}^h, \qquad i = 1, \ldots, N. \tag{23}$$

State-space representation. Averaging Equation (23) over the cross-sectional dimension i and solving for π_t yields

$$\pi_t = \delta^h + \mu_t^h + \beta_t^h \, \widehat{\pi}_{t\,|\,t-h}^{\text{av}} + \theta_t^h + \varepsilon_t^h, \tag{24}$$

By defining $\alpha_t^h = \delta^h + \mu_t^h$ and $u_t^h = \theta_t^h + \varepsilon_t^h$, for $t = 1, \ldots, T$, we can write a state-space model given by

$$\begin{aligned}
\pi_t &= \alpha_t^h + \beta_t^h \, \widehat{\pi}_{t\,|\,t-h}^{\text{av}} + u_t^h \\
\alpha_t^h &= \alpha_{t-h}^h + \nu_t^h \\
\beta_t^h &= \beta_{t-h}^h + \eta_t^h \\
\begin{pmatrix} u_t^h \\ \nu_t^h \\ \eta_t^h \end{pmatrix} &\sim \mathcal{N}\left(\mathbf{0},\ \mathsf{diag}\left(\sigma_u^2,\ \sigma_\nu^2,\ \sigma_\eta^2\right)\right),
\end{aligned}$$

where, likewise as before, we assume a multivariate normal distribution for (independent) disturbances u_t^h, ν_t^h, and η_t^h.

Time-varying EBCAF (TV-EBCAF). Just like before, considering a sample for $t = 1, \ldots, T$, we estimate $\{\alpha_t^h, \beta_t^h\}_t$ by maximum likelihood using the Kalman filter recursion. Thus, a version of time-varying extended bias-corrected average forecast is given by

$$\widehat{\pi}_{T+h\,|\,T}^{\text{TV-EBCAF}} = \widehat{\alpha}_{T+h\,|\,T}^h + \widehat{\beta}_{T+h\,|\,T}^h \, \widehat{\pi}_{T+h\,|\,T}^{\text{av}},$$

where $\widehat{\alpha}_{T+h\,|\,T}^h$ and $\widehat{\beta}_{T+h\,|\,T}^h$ are *predict values* for state variables α_{T+h}^h and β_{T+h}^h, respectively, both recovered using the Kalman filter.

3.3 Estimation

Just like before, we set January 2015 as the starting point for generating combined forecasts for all horizons. The sample for estimation begins in January 2012. To ensure the positivity of the error term variances, we impose an exponential form given by $\sigma^2 = \mathsf{exp}(\tau)$, where τ is a parameter to be estimated by maximum likelihood. In the TV-BCAF and TV-EBCAF models, there are two and three parameters to assess, respectively. To initialize the maximum likelihood estimator, we consider the OLS estimates generated from linear models with time-fixed parameters and the first 36 available observations, when applicable.[2] Thus, the initial values of the time-varying parameters (TPVs) are the OLS estimates for the coefficients of this initial stage. The starting variance-covariance matrix of TVPs corresponds to the conventional estimator for the variance-covariance matrix of parameters in the initial model. For the variance of error terms of the measurement equation, we use the variance of residuals from the initial model estimated by OLS. Finally, we set the initial variance of a state equation error to be $0.2^2 = 0.04$. We employ a quasi-Newton method (BFGS) as the optimization algorithm.

4 Results and discussion

Table 2 exhibits results of survey-based expectations, average forecast, BCAF based on both 9-month rolling windows and expanding windows, EBCAF based on expanding windows, and TV-BCAF and TV-EBCAF models for the period from January 2015 to June 2022. Figure 1 shows the evolution over time and by forecast horizon of actual inflation and forecasts resulting from some combination of forecasts considering the set of individual forecasts for the aggregate inflation (direct forecasting approach), including the available Focus consensus.

Firstly, it can be observed that the TV-BCAF models, which only include time-varying intercept bias (no slope bias), do not perform well with respect to their counterparts based on 9-month rolling windows or expanding windows and average forecast: almost all RMSE ratios are higher across forecast horizons and individual forecast sets addressed. Looking at the evolution of the forecasts obtained by the TV-BCAF, these models basically extrapolate the present forecast error into the future. The forecast error tends to propagate farther into the future as the forecast horizon extends.

For shorter and intermediate horizons, the inclusion of the time-varying slope bias correction (TV-EBCAF model) produces improvements compared to the model that solely incorporates a time-varying intercept bias. Although the TV-EBCAF model's numerical performance falls slightly behind the model that corrects only for intercept bias using a 9-month estimation window, there are marginal improvements when compared to the EBCAF model based on extended windows. This outcome suggests that a full correction for time-varying biases can be advantageous for inflation forecasting. When comparing the forecasts generated by TV-BCAF and TV-EBCAF models, we observe that the latter carries fewer ex-post errors forward, except for a noticeable overprediction of inflation in 2016 and 2017. It is worth noting that for more distant horizons, the prominent positive bias, in early 2019, and negative bias, in early 2020, are mitigated to some extent in the case of the TV-EBCAF model

[2] As explained in the Footnote 1, for early periods, we may have fewer than 36 observations for some horizons. For example, in Jan/2015, only 25 stayed available for horizon $h = 12$. However, starting from Jan/2016, there are 36 starting observations available for all horizons.

Table 2: Out-of-sample RMSE ratio for average forecast and time-varying BCAF and EBCAF models with respect to available Focus: Jan/2015 to Jun/2022

Method/Forecasts	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$	$h=11$
A. Survey												
Focus (available)	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*1.000*	*1.000*
Focus (*ex-post*)	**0.931*****	*0.972****	0.993***	1.001	1.000	1.000	1.000	0.999	0.999*	1.000	*1.001*	*1.001*
B. Average forecast												
Aggregates & Focus	1.215	*0.989*	*0.953***	*0.971**	*0.982*	*0.970*	**0.939****	**0.928*****	**0.930****	**0.935****	**0.968***	**0.986**
BCB & Focus	1.208	1.004	0.967	0.989	0.994	0.977	*0.945***	*0.951**	*0.948***	**0.950****	**0.973**	**0.985**
Groups & Focus	1.341	1.049	0.971	0.991	*0.984*	*0.968*	0.968	*0.950**	*0.953**	*0.976*	*0.975*	*0.992*
Subgroups & Focus	1.514	1.110	1.026	1.031	1.019	0.998	0.973	*0.960*	0.977	*0.996*	1.016	1.026
All & Focus	1.317	1.033	0.974	0.991	0.990	0.972	*0.949***	**0.940****	**0.945***	*0.957**	*0.977*	*0.993*
C. Bias-corrected average forecast (BCAF) with rolling windows of 9 months												
Aggregates & Focus	1.209	**0.958**	**0.904***	**0.924**	**0.955**	*0.958*	**0.945**	0.964	0.993	1.039	1.093	1.127
BCB & Focus	1.239	*0.979*	**0.907***	**0.936**	*0.961*	*0.966*	0.955	0.991	1.015	1.052	1.103	1.116
Groups & Focus	1.355	1.030	*0.925*	*0.955*	**0.954**	**0.949**	0.961	0.983	1.019	1.097	1.120	1.140
Subgroups & Focus	1.500	1.074	0.972	*0.973*	*0.973*	0.972	0.963	0.985	1.047	1.113	1.163	1.185
All & Focus	1.327	1.011	*0.925*	*0.945*	*0.957*	*0.958*	*0.951*	0.978	1.016	1.074	1.118	1.141
D. Bias-corrected average forecast (BCAF) with expanding windows												
Aggregates & Focus	1.217	*0.987*	*0.954**	*0.972*	0.986	0.976	*0.953**	*0.948**	*0.957*	*0.969*	*1.005*	*1.025*
BCB & Focus	1.216	1.008	0.974	1.001	1.010	0.996	0.971	0.982	0.981	*0.986*	*1.012*	*1.024*
Groups & Focus	1.350	1.057	0.979	1.005	1.002	0.990	0.995	0.982	0.992	1.020	1.020	1.034
Subgroups & Focus	1.520	1.112	1.033	1.045	1.041	1.024	1.003	0.996	1.018	1.045	1.070	1.079
All & Focus	1.324	1.038	0.981	1.003	1.007	0.993	0.975	0.972	0.982	1.000	1.023	1.038
E. Extended bias-corrected average forecast (EBCAF) with expanding windows												
Aggregates & Focus	*1.074*	**0.967**	*0.967*	0.987	1.013	1.009	0.996	0.998	1.016	1.028	1.039	1.090
BCB & Focus	*1.139*	*0.990*	0.986	1.013	1.026	1.007	0.984	0.998	0.997	*0.996*	1.017	1.046
Groups & Focus	1.208	1.046	0.987	1.019	1.019	1.005	1.009	1.005	1.009	1.028	1.025	1.034
Subgroups & Focus	1.302	1.120	1.049	1.045	1.043	1.030	1.002	0.996	1.011	1.031	1.069	1.056
All & Focus	*1.153*	1.023	0.995	1.017	1.027	1.013	0.998	1.004	1.011	1.026	1.046	1.069
F. Time-varying bias-corrected average forecast (TV-BCAF)												
Aggregates & Focus	1.233	1.005	1.018	1.040	1.070	1.049	0.995	0.989	1.001	1.052	1.121	1.184
BCB & Focus	1.226	1.026	1.005	1.034	1.040	1.018	*0.953*	*0.959*	*0.975*	1.030	1.104	1.134
Groups & Focus	1.364	1.103	1.044	1.079	1.067	1.019	1.004	0.991	0.975	1.056	1.110	1.149
Subgroups & Focus	1.512	1.157	1.122	1.114	1.096	1.025	0.993	0.993	0.996	1.068	1.142	1.194
All & Focus	1.336	1.079	1.062	1.072	1.071	1.026	0.987	0.987	0.986	1.049	1.120	1.166
G. Time-varying extended bias-corrected average forecast (TV-EBCAF)												
Aggregates & Focus	*1.087*	*0.998*	*0.948*	*0.973*	1.009	0.977	0.957	0.970	1.016	1.053	1.084	1.146
BCB & Focus	*1.145*	1.000	0.972	0.984	1.018	1.013	*0.947*	*0.960*	0.981	1.019	1.099	1.123
Groups & Focus	1.211	1.086	0.985	1.024	0.993	*0.964*	0.970	0.968	*0.969*	1.018	1.088	1.133
Subgroups & Focus	1.330	1.135	1.044	1.087	1.066	**0.952**	0.997	1.000	*0.964*	1.013	1.100	1.145
All & Focus	*1.156*	1.046	0.986	0.989	*0.979*	0.971	0.956	0.973	0.988	1.037	1.089	1.163

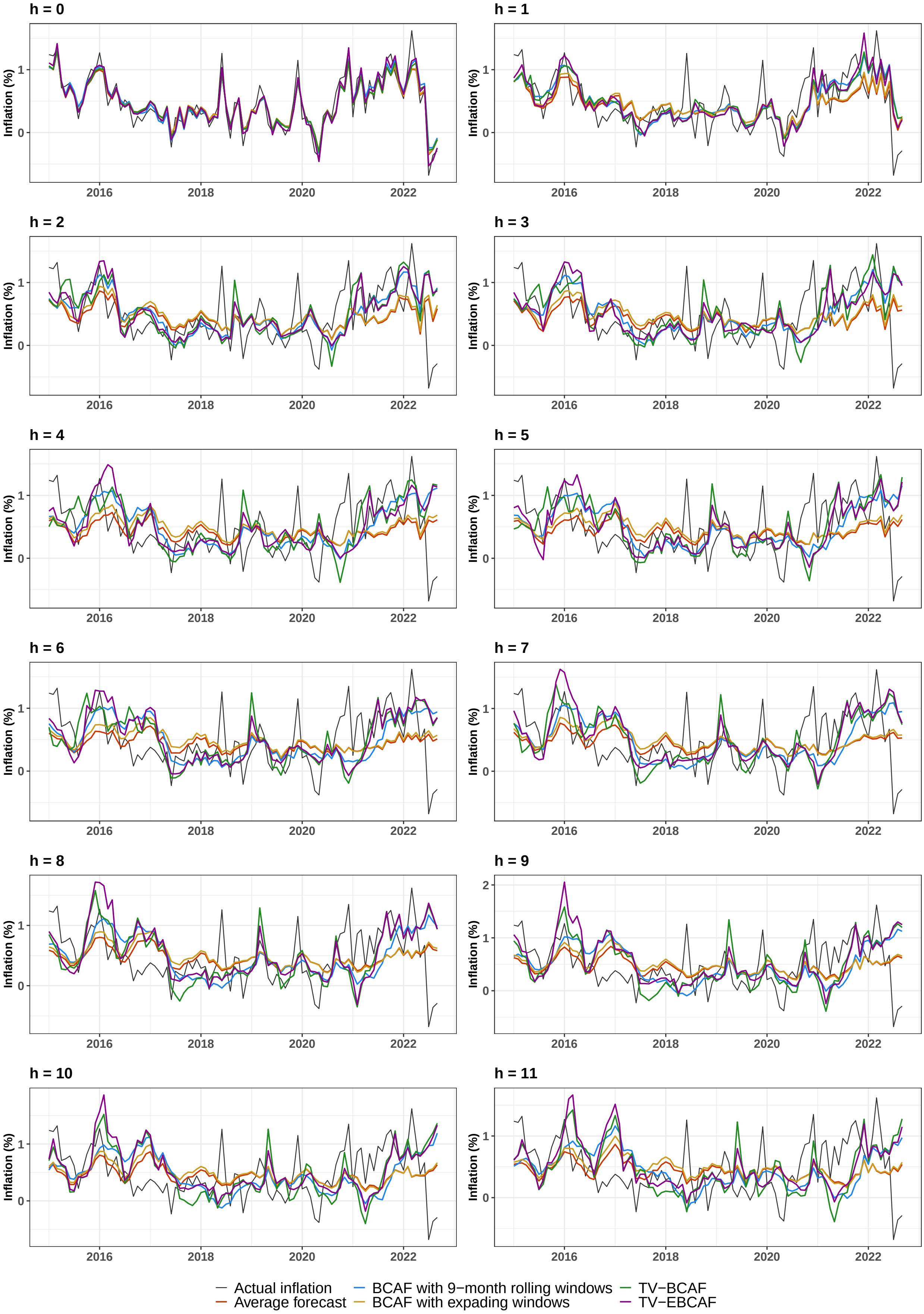
Notes: see Table 1.

compared to the TV-BCAF model. This highlights the importance of incorporating a slope bias in the time-varying bias-correction approach.

By assuming a random walk process for the time-varying bias, we expect the forward loading of ex-post forecast errors. In this regard, we are investigating alternative specifications that incorporate an autoregressive (AR) model with a non-zero intercept and an AR term smaller than one for the states. However, the results obtained so far are not satisfactory. We are considering specifications that deviate from the traditional local-level model or assumptions regarding the random walk process for latent variables. It is important to preserve parsimony in specifying a state-space model since the estimation of these models may suffer from severe instabilities, along with difficulties associated with identification.

Finally, an important finding of our essay is the lack of statistical superiority of most forecasts generated by bias correction models. Although, on average, some models present

Figure 1: Forecasts generated by selected combinations considering the set of individual forecasts for aggregate directly, by horizon



a lower RMSE ratio than the average forecast, traditional models with fixed biases or models based on time-varying bias hardly demonstrate statistically superior performance compared to the Focus consensus. This outcome may be attributed to various factors. Firstly, the improved performance of bias-corrected procedures may be closely linked to the period of the COVID-19 pandemic, characterized by higher and more volatile inflation. Secondly, the uncertainty associated with parameter estimation in the less parsimonious corrected-bias models could offer another potential explanation. More specifically, models incorporating time-varying bias display a notable divergence of estimators for these latent variables, as evidenced by the variance-covariance matrix of the states. In practical terms, this instability in estimation over time leads to the occurrence of atypical forecasts, thereby hindering the achievement of statistical significance.

5 Final considerations

In this essay, we introduce a model for correcting the bias in the average forecast, which allows both intercept and slope biases to vary over time. Initially, we proceed with an estimation based on rolling windows of different lengths. Such models allow the parameters to oscillate over time to a greater degree than the variation that occurs in a procedure based on extended windows. Applying the procedure to different sets of individual inflation forecasts, we find good predictive results for these rolling-window-based models, particularly for windows ranging from 6 to 12 months in the case of intermediate forecast horizons (one to six months ahead). Based on this result, we suggest a state-space model that allows for obtaining time-varying bias components using all available information, that is, without the need to define *ad-hoc* a window size. Overall, the model that includes corrections for intercept and slope bias varying over time tends to perform slightly worse than rolling-window-based procedures. However, it is worth investigating other specifications for the state-space model and alternatives for reducing the variance of the estimated time-varying biases.

Appendix A Kalman filter

Following Hamilton (1994, Chapter 13) and Elliott and Timmermann (2016, Appendix A), let $\boldsymbol{y}_t$ be a n-dimensional vector of (observable) variable observed at period t, and $\boldsymbol{\xi}$ be a r-dimensional vector of state (unobservable) variables. Consider a generic state-space model consisting of a measurement equation, a state equation, and a perturbation equation as follows:

$$\boldsymbol{y}_t = \boldsymbol{H}\,\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\xi}_t = \boldsymbol{F}\,\boldsymbol{\xi}_{t-h} + \boldsymbol{\nu}_t$$

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\nu}_t \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P} \end{bmatrix}\right),$$

where $\boldsymbol{H}$ and $\boldsymbol{F}$ are matrices of parameters, $\boldsymbol{Q}$ and $\boldsymbol{P}$ are covariance matrices, and $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\nu}_t$ are independent vectors of white noise. Given starting values $\widehat{\boldsymbol{\xi}}_{h\,|\,0}$ and $\boldsymbol{P}_{h\,|\,0}$, we recursively compute the following values until we get to $\widehat{\boldsymbol{\xi}}_{t+h\,|\,t}$ and $\boldsymbol{P}_{t+h\,|\,t}$, and then compute $\widehat{\boldsymbol{y}}_{t+h\,|\,t}$. We estimate the parameters in the matrices $\boldsymbol{F}$, $\boldsymbol{H}$, $\boldsymbol{Q}$, and $\boldsymbol{P}$ employing the maximum likelihood estimator combined with the recursiveness of the Kalman filter.

Appendix B Forecast error and estimated intercept bias

Figure B.1: Forecast error and estimate intercept bias over time: from $h = 0$ to $h = 3$, by set of forecasts

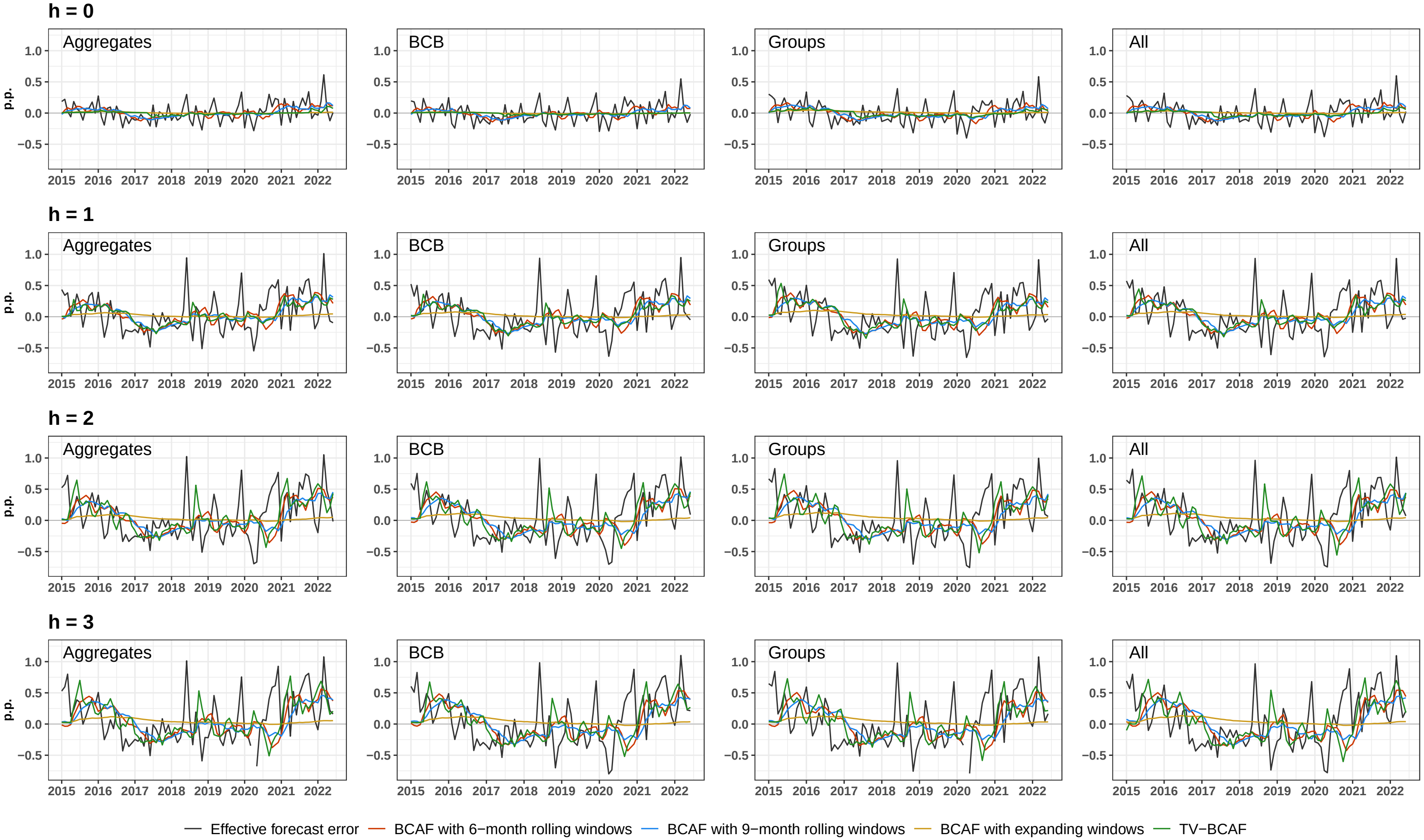


Figure B.2: Forecast error and estimate intercept bias over time: from $h = 4$ to $h = 7$, by set of forecasts

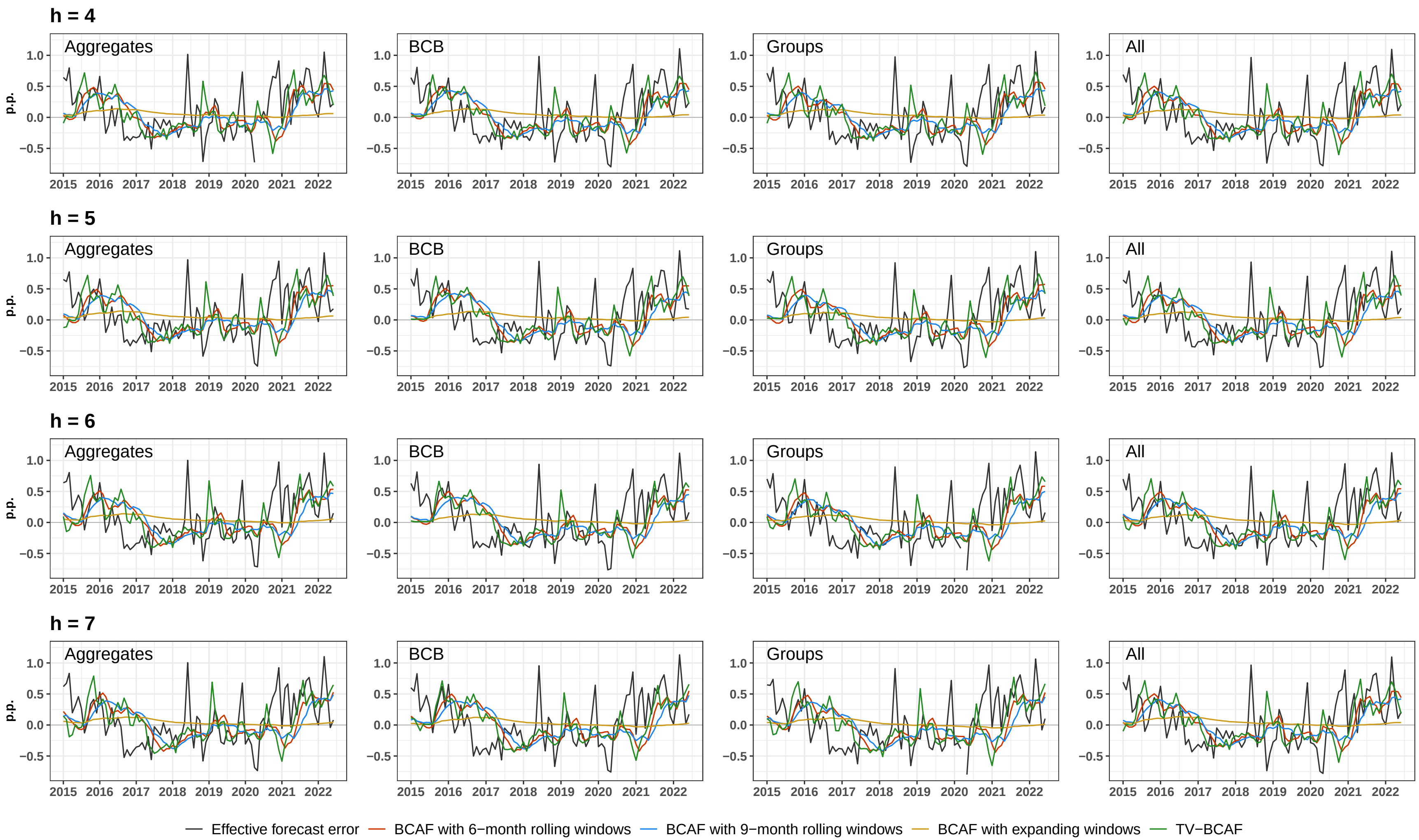
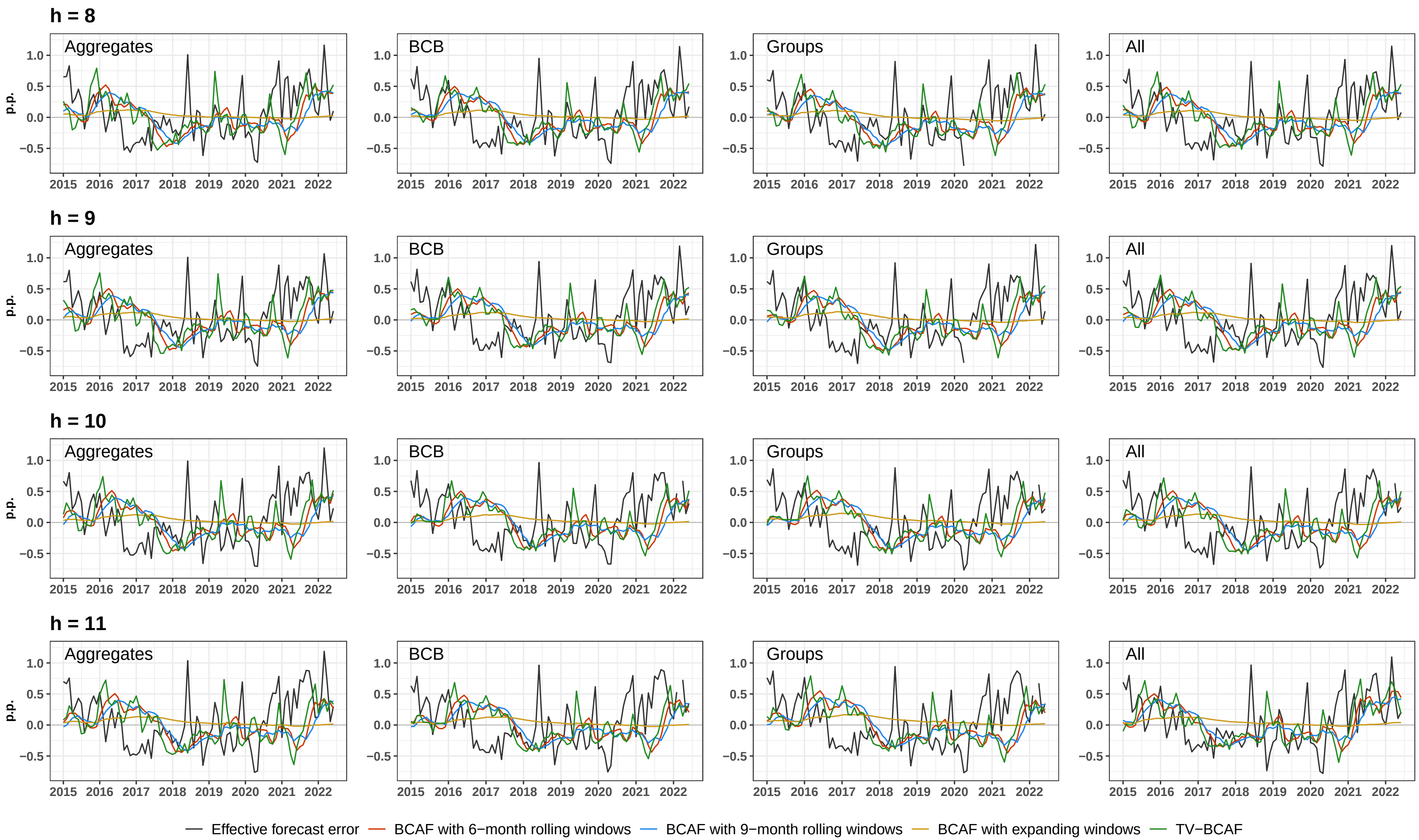


Figure B.3: Forecast error and estimate intercept bias over time: from $h = 8$ to $h = 11$, by set of forecasts



References

BATES, J. M., AND C. W. GRANGER (1969): "The Combination of Forecasts," *Journal of the Operational Research Society*, 20(4), 451–468.

CAPISTRÁN, C., AND A. TIMMERMANN (2009): "Forecast combination with entry and exit of experts," *Journal of Business & Economic Statistics*, 27(4), 428–440.

CARVALHO, F. A., AND A. MINELLA (2012): "Survey forecasts in Brazil: A prismatic assessment of epidemiology, performance, and determinants," *Journal of International Money and Finance*, 31(6), 1371–1391.

ELLIOTT, G., AND A. TIMMERMANN (2016): *Economic Forecasting*. Princeton University Press, Princeton, New Jersey.

GAGLIANONE, W. P., AND J. V. ISSLER (2023): "Microfounded Forecasting," Mimeo., FGV EPGE.

GAGLIANONE, W. P., J. V. ISSLER, AND S. M. MATOS (2017): "Applying a microfounded-forecasting approach to predict Brazilian inflation," *Empirical Economics*, 53(1), 137–163.

GRANGER, C. W., AND R. RAMANATHAN (1984): "Improved methods of combining forecasts," *Journal of forecasting*, 3(2), 197–204.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.

ISSLER, J. V., AND L. R. LIMA (2009): "A panel data approach to economic forecasting: The bias-corrected average forecast," *Journal of Econometrics*, 152(2), 153–164.

MINCER, J. A., AND V. ZARNOWITZ (1969): "The Evaluation of Economic Forecasts," in *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, pp. 3–46. NBER.

NEWBOLD, P., AND C. W. GRANGER (1974): "Experience with forecasting univariate time series and the combination of forecasts," *Journal of the Royal Statistical Society: Series A (General)*, 137(2), 131–146.

PALM, F. C., AND A. ZELLNER (1992): "To combine or not to combine? Issues of combining forecasts," *Journal of Forecasting*, 11(8), 687–701.

REID, D. J. (1969): "A comparative study of time series prediction techniques on economic data," Ph.D. thesis, Dept. of Mathematics, University of Nottingham.