CrossMark

Applying a microfounded-forecasting approach to predict Brazilian inflation

Wagner Piazza Gaglianone[1] · João Victor Issler[2] ·
Silvia Maria Matos[3]

Abstract We investigate whether combining forecasts from surveys of expectations is a helpful empirical strategy for forecasting inflation in Brazil. We employ the FGV–IBRE Economic Tendency Survey, which consists of monthly qualitative information from approximately 2000 consumers since 2006, and also the Focus Survey of the Central Bank of Brazil, with daily forecasts since 1999 from roughly 250 professional forecasters. Natural candidates to win a forecast competition in the literature of surveys of expectations are the (consensus) cross-sectional average forecasts (AF).

✉ João Victor Issler
    jissler@fgv.br

    Wagner Piazza Gaglianone
    wagner.gaglianone@bcb.gov.br

    Silvia Maria Matos
    silvia.matos@fgv.br

[1] Research Department, Central Bank of Brazil, Rio de Janeiro, Brazil

[2] Graduate School of Economics, Getulio Vargas Foundation, Praia de Botafogo 190, s.1104,
    Rio de Janeiro, RJ 22.253-900, Brazil

[3] Instituto Brasileiro de Economia – IBRE, Getulio Vargas Foundation, Rio de Janeiro, Brazil

We first show that these forecasts are a bias-ridden version of the conditional expectation of inflation using the no-bias tests proposed in Issler and Lima (J Econom 152(2):153–164, 2009) and Gaglianone and Issler (Microfounded forecasting, 2015). The results reveal interesting data features: Consumers systematically overestimate inflation (by 2.01 p.p., on average), whereas market agents underestimate it (by 0.68 p.p. over the same sample). Next, we employ a pseudo out-of-sample analysis to evaluate different forecasting methods: the AR(1) model, the Granger and Ramanathan (J Forecast 3:197–204, 1984) forecast combination (GR) technique, a Phillips-curve based method, the Capistrán and Timmermann (J Bus Econ Stat 27:428–440, 2009) combination method, the consensus forecast (AF), the bias-corrected average forecast (BCAF), and the extended BCAF. Results reveal that: (i) the MSE of the AR(1) model is higher compared to the GR (and usually lower compared to the AF); and (ii) the extended BCAF is more accurate than the BCAF, which, in turn, dominates the AF. This validates the view that the bias corrections are a useful device for forecasting using surveys. The Phillips-curve based method has a median performance in terms of MSE, and the Capistrán and Timmermann (2009) combination method fares slightly worse.

Keywords Consensus forecasts · Forecast combination · Common features · Panel data

JEL Classification C14 · C33 · E37

1 Introduction

From a theoretical and empirical point-of-view, Bates and Granger (1969) made the econometric profession aware of the benefits of forecast combination. Their empirical results were later confirmed by a variety of time-series studies—e.g., Granger and Ramanathan (1984), Palm and Zellner (1992), Stock and Watson (2002a, b, 2006), Timmermann (2006), and Genre et al. (2013)—where the simple average of forecasts is shown to perform relatively well.

In a pioneering study on forecasting using panel-data techniques, Davies and Lahiri (1995) proposed testing forecast rationality using information across forecasts and forecast horizons by employing a three-way decomposition.[1] This lead to a subsequent literature on what could be labeled *a panel-data approach to forecasting*; see Baltagi (2013) for a broader discussion on this topic. Some of these studies focused on forecast combination and others on density-forecast estimation, e.g., Issler and Lima (2009), Gaglianone et al. (2011), Gaglianone and Lima (2012, 2014), Gaglianone and Issler (2015), and Lahiri et al. (2015).

In Issler and Lima and Gaglianone and Issler, the main idea is that the *consensus forecast* (the cross-sectional average of individual forecasts) is a bias-ridden version

---

[1] Prior to that, Palm and Zellner (1992) also employ a two-way decomposition to discuss forecast combination in a Bayesian and a non-Bayesian setup. But, the main focus of their paper is not on panel-data techniques, but on the usefulness of a Bayesian approach to forecasting.

of the conditional expectation, which is the optimal forecast under a mean-squared-error (MSE) *risk function*. Once these potential biases are properly removed, one can construct forecasting devices that are optimal in the limit, i.e., that can approach the conditional expectation asymptotically.

In this paper, we show empirically that combining forecasts in surveys is a promising empirical strategy to increase forecast accuracy. Our focus is on Brazilian inflation, a key variable on the *Inflation Targeting Regime* , implemented by the Central Bank of Brazil (BCB). In Brazil, there are two main surveys of expectations regarding inflation. The first is the *Focus Survey*—a survey of professional forecasters—kept by the BCB. The second is the *FGV–IBRE Economic Tendency Survey*—a survey of individual consumers. Both include forecasts for the official index used in the Inflation Targeting Regime—the IPCA, a Broad Consumer Price Index (CPI). Using information in these surveys, we employ the techniques in Issler and Lima and in Gaglianone and Issler to construct potentially optimal forecasts under MSE risk. The method discussed in the former shows that the mean of the consensus forecast is biased. The method discussed in the latter is *microfounded* and shows that the optimal forecast decision faced by an individual agent is related to the conditional expectation by an *affine* structure. This leads to the fact that the consensus forecast has two sources of bias—an intercept bias and a slope bias. Fortunately, they can be estimated to filter out the conditional expectation from a database of survey responses.

Using bias correction for the consensus forecast is a relevant empirical issue, since there is a growing forecasting literature in macroeconomics where the consensus forecast is used with no correction and rationality tests employing it usually find a result opposite to rationality; see, e.g., Coibion and Gorodnichenko (2012). Below, we show some evidence of these phenomena and discuss some important implications regarding rationality tests.

The usefulness of survey data for forecasting has been recently shown in a variety of studies. For example, Ang et al. (2007) find that true out-of-sample survey forecasts (e.g., Michigan; Livingston) outperform a large number of out-of-sample single-equation and multivariate time-series competitors. Faust and Wright (2013) argue that subjective forecasts of inflation based on surveys seem to outperform model-based forecasts in certain dimensions, often by a wide margin. In a world where there is an increasing availability of reliable survey data provided electronically, it is interesting to examine how one could efficiently use it. This is exactly the objective of this paper, where we exploit the fact that the Focus Survey has information on more than 250 professional forecasters and that the Economic Tendency Survey keeps responses of about 1500 respondents on inflation expectations in every month.

The rest of the paper is divided as follows. Section 2 discusses the econometric techniques employed in this paper, with some focus on the work of Issler and Lima (2009) and of Gaglianone and Issler (2015), but covering the adjacent literature as well. Section 3 presents a real-time forecasting exercise with data from the FGV survey of consumer expectations on inflation and from the Focus survey of the BCB on inflation expectations of professional forecasters. It also discusses rationality tests using the consensus forecasts and its possible shortcomings. Section 4 concludes.

2 Econometric methodology

Forecast combination has proved to be a valuable tool at least since Bates and Granger (1969). Palm and Zellner (1992) and Davies and Lahiri (1995) pioneered the use of panel-data techniques in forecasting. This section discusses in some detail the forecast-combination approach put forth by Gaglianone and Issler (2015) on how to combine survey expectations to obtain optimal forecasts in a panel-data context. Some parts of the material therein can also be found in Issler and Lima (2009) and in Lahiri et al. (2015).

The techniques discussed in this section are appropriate for forecasting a weakly stationary and ergodic univariate process $\{y_t\}$ using a large number of forecasts, coming from forecast surveys (expectations) on the variable in question – y_t. Some (or all) of these responses can be generated by using econometric models, but then we have no knowledge of them. We label individual forecasts of y_t, computed using information sets lagged h periods, by $f_{i,t}^h, i = 1, 2, \ldots, N, h = 1, 2, \ldots, H$, and $t = 1, 2, \ldots, T$. Therefore, $f_{i,t}^h$ are h-step-ahead forecasts of y_t, formed at period $t - h$, and N is the number of respondents of this opinion poll regarding y_t. Gaglianone and Issler show that, in a variety of interesting cases, optimal forecasts are related to $\mathbb{E}_{t-h}(y_t)$—the conditional expectation of y_t, computed using information lagged h periods—by an affine function:

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

This result is somewhat expected given the pioneering work of Granger (1969). He shows that, under a mean-squared-error (MSE) *risk function* and proper regularity conditions, the optimal forecast is equal to $\mathbb{E}_{t-h}(y_t)$. Moreover, if the risk function is asymmetric and proper regularity conditions are met, then Granger showed that the optimal forecast is equal to $\mathbb{E}_{t-h}(y_t) + k_i^h$; see also the later developments in Christof-fersen and Diebold (1997), Elliott and Timmermann (2004), Patton and Timmermann (2007), and Elliott et al. (2008), for example.

Gaglianone and Issler consider a setup with two layers of decision making. In the first layer, individuals (survey respondents) form their *optimal* point forecasts $\left(f_{i,t}^h\right)$ of a random variable y_t by using a specific *loss function*. They allow for the existence of asymmetry of the loss function and different assumptions about knowledge of the DGP of y_t. The optimal forecasts $f_{i,t}^h$ will be available as survey results, where the number of respondents is potentially large, i.e., $N \to \infty$, and these surveys can be periodically taken on a large number of different occasions, i.e., $T \to \infty$. In the second layer of decision making, an econometrician will be the final user of this large number of forecasts, operating under an MSE *risk function*. Her/his challenge is to uncover $\mathbb{E}_{t-h}(y_t)$—the optimal forecast of the second layer of decision making.

Gaglianone and Issler employ the basic assumptions in the econometric forecast literature to derive a consistent estimate of $\mathbb{E}_{t-h}(y_t)$, which can be obtained by using Hansen's (1982) generalized method of moments (GMM). Start by taking a cross-sectional average of (1), after using:

$$y_t = \mathbb{E}_{t-h}(y_t) - \eta_t^h, \tag{2}$$

noting that η_t^h is a martingale-difference sequence by construction, i.e., $\mathbb{E}_{t-h}(\eta_t^h) = 0$. Under suitable assumptions, they show that the following system of equations can be used to estimate the key parameters $\overline{k^h}$ and $\overline{\beta^h}$ in:

$$\mathbb{E}\left[\left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \otimes z_{t-s}\right] = 0, \quad h = 1, 2, \ldots, H, \tag{3}$$

where z_{t-s} is a vector of instruments, dated $t-s$ or older, $s \geq h$, $\overline{f_{\cdot,t}^h} = \frac{1}{N}\sum_{i=1}^N f_{i,t}^h$, $\overline{k^h} = \frac{1}{N}\sum_{i=1}^N k_i^h$ and $\overline{\beta^h} = \frac{1}{N}\sum_{i=1}^N \beta_i^h$, for each h. Over-identification requires that $dim(z_{t-s}) > 2$.

Gaglianone and Issler argue that there is no need to estimate individual coefficients k_i^h and β_i^h—only their means to be able to identify and estimate $\mathbb{E}_{t-h}(y_t)$. In doing so, they avoid the curse of dimensionality given that $N \to \infty$. As long as these cross-sectional averages converge, GMM using time-series restrictions delivers consistent estimates of the respective parameter means.

Current surveys, however, usually approximate better the case where $T \to \infty$, while N is small or diverges at a smaller rate than T. Under additional conditions on the cross-sectional averages $\overline{f_{\cdot,t}^h} = \frac{1}{N}\sum_{i=1}^N f_{i,t}^h, \overline{k^h} = \frac{1}{N}\sum_{i=1}^N k_i^h$ and $\overline{\beta^h} = \frac{1}{N}\sum_{i=1}^N \beta_i^h$, Gaglianone and Issler show that one can still estimate $\mathbb{E}_{t-h}(y_t)$ consistently, when $T \to \infty$ first, and then $N \to \infty$.

One way to exploit all possible moment conditions implicit in (3) is to stack all the restrictions across h (finite) as:

$$\mathbb{E}\left[\begin{pmatrix} \overline{f_{\cdot,t}^1} - \overline{k^1} - \overline{\beta^1} y_t \\ \overline{f_{\cdot,t}^2} - \overline{k^2} - \overline{\beta^2} y_t \\ \vdots \\ \overline{f_{\cdot,t}^H} - \overline{k^H} - \overline{\beta^H} y_t \end{pmatrix} \otimes z_{t-s}\right] = 0, \tag{4}$$

where the problem collapses to one where we have $H \times dim(z_{t-s})$ restrictions and $2H$ parameters to estimate. As before, over-identification requires that $dim(z_{t-s}) > 2$. Given a choice of H, GMM estimation of (4) is efficient.[2] A less efficient alternative to estimate the whole stacked system (4) is to estimate separately (3) for every horizon

---

[2] One could impose additional restrictions across coefficients $\overline{k^h}$ and $\overline{\beta^h}$ using the term structure of forecasts. Given stationarity of y_t: $y_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \cdots + \psi_{h-1}\varepsilon_{t-h+1} + \psi_h\varepsilon_{t-h} + \cdots$, which gives: $\mathbb{E}_{t-(h-1)}(y_t) = \mathbb{E}_{t-h}(y_t) + \psi_{h-1}\varepsilon_{t-(h-1)}$, $h = 1, 2, \ldots, H$, showing that the expectation revision $\mathbb{E}_{t-(h-1)}(y_t) - \mathbb{E}_{t-h}(y_t)$ is unforecastable using information up to $t-h$. In a GMM context, this implies additional moment restrictions across horizons if we substitute $\mathbb{E}_{t-h}(y_t)$ by $\frac{1}{N}\sum_{i=1}^N \frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}}$ and $\mathbb{E}_{t-(h-1)}(y_t)$ by $\frac{1}{N}\sum_{i=1}^N \frac{f_{i,t}^{h-1} - \overline{k^{h-1}}}{\overline{\beta^{h-1}}}$. Notice that, for large enough N, we can approximate reasonably well these conditional expectations using these averages, as proposed in Gaglianone–Issler. They did not consider these additional restrictions into GMM estimation, but that could certainly be exploited in the future.

h, which could also be attempted for computational reasons. As is usual within GMM estimation, one can perform a standard misspecification test of correlation between errors and instruments by employing the J-test proposed by Hansen (1982).

Using GMM estimates, $\widehat{\overline{k^h}}$ and $\widehat{\overline{\beta^h}}$, obtained when $T \to \infty$, Gaglianone and Issler show the critical result that:

$$\frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}} \xrightarrow{p} \mathbb{E}_{t-h}(y_t), \tag{5}$$

in a variety of different setups: when $T \to \infty$ first, and then $N \to \infty$; when $N \to \infty$ first, and then $T \to \infty$; and when $N, T \to \infty$ in a joint limit framework, all using the method in Phillips and Moon (1999). The conditions under which (5) holds vary according to the asymptotic setup.

Equation (5) offers a way to filter the consensus forecast $\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h$ extracting $\mathbb{E}_{t-h}(y_t)$ in the limit. Thus, it delivers an estimate of the optimal forecast in the second layer of decision making. It can be viewed immediately as a forecast-combination method.

A paper that delivers a similar result, although suggests a different estimation method, is Capistrán and Timmermann (2009). Start with Eq. (1) and take the cross-sectional average to obtain[3]:

$$\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h = \frac{1}{N} \sum_{i=1}^{N} \left(k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h \right), \quad \text{or, in standard notation,}$$
$$\overline{f_{\cdot,t}^h} = \overline{k^h} + \overline{\beta^h} \mathbb{E}_{t-h}(y_t) + \overline{\varepsilon_{\cdot,t}^h} \tag{6}$$

Combine (6) with (2) to obtain, for large enough N:

$$y_t = -\frac{\overline{k^h}}{\overline{\beta^h}} + \frac{1}{\overline{\beta^h}} \overline{f_{\cdot,t}^h} - \eta_t^h. \tag{7}$$

The approach in Capistrán and Timmermann yields a single equation where $\overline{k^h}$ and $\overline{\beta^h}$ can be estimated to filter $\mathbb{E}_{t-h}(y_t)$ in a similar fashion to that in (5). As can be seen by comparing (7) and (3), these two equations identify $\overline{k^h}$ and $\overline{\beta^h}$. A potential advantage for the method in Gaglianone and Issler using (4) is that it allows stacking moment conditions across horizons and joint estimation of the bias terms. Thus, it benefits from the efficiency gains associated with employing a multi-equation setup as opposed to the single-equation setup in Capistrán and Timmermann. In the empirical section, we shall compare directly these two methods regarding their out-of-sample mean-squared error.

---

[3] We thank an anonymous referee for pointing this out to us.

In an out-of-sample forecasting exercise involving surveys, there are usually three consecutive distinct time sub-periods, where time is indexed by $t = 1, 2, \ldots, T_1, \ldots, T_2, \ldots, T$. The first sub-period E is labeled as 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 $(t = 1, 2, \ldots, T_1)$. For the other two, we follow the standard notation in West (1996). The sub-period R (for regression) is labeled as 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 $(t = T_1 + 1, \ldots, T_2)$. 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 $(t = T_2 + 1, \ldots, T)$. 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 $O(T)$.

It is worth mentioning that the method proposed by Gaglianone and Issler extends the previous literature of forecasting within a panel-data framework, e.g., Palm and Zellner (1992), Davies and Lahiri (1995), Issler and Lima (2009), and Lahiri et al. (2015).

For example, their setup encompasses that of Davies and Lahiri (1995), reproduced below with our notation:

$$y_t - f_{i,t}^h = - \left(k_i^h + \eta_t^h + \varepsilon_{i,t}^h \right), \tag{8}$$

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

The method proposed by Gaglianone and Issler is appropriate to cover any type of survey where potentially the number of observations T is large, encompassing the cases where the number of survey respondents and of time observations is large—*big data*, and also the case where the number of time observations is large but the number of respondents is fixed—standard continuous macroeconomic surveys.

The way Gaglianone and Issler identify the conditional expectation can be viewed as a combination of cross-sectional averages with standard GMM moment restrictions, where the affine structure offers natural orthogonality restrictions allowing the estimation of bias-correction terms. They circumvent the *curse of dimensionality* that arises from the factor structure (large N) by employing these cross-sectional averages.

3 Empirical application

3.1 Data

3.1.1 FGV's consumer survey

The FGV–IBRE Economic Tendency Survey compiles business and consumers expectations of key economic series in Brazil. The Brazilian Institute of Economics (IBRE) is a pioneer in surveys, and this one runs since 1966. Since September, 2005, FGV–IBRE conducts a monthly consumer survey, which consists of qualitative information on household consumption, savings, financial variables, employment, etc. The survey has a country-wide coverage (seven major state capitals) with approximately 2000 consumers. Survey respondents are classified into four classes of household monthly income as follows[4]: *Income Level 1—Up to R$2100; Income Level 2—Between R$ 2100.01 and   R$ 4800.00; Income Level 3—Between R$4800.01 and R$ 9600.00; Income Level 4—More than R$ 9600.01.* Survey information can also be broken down by different classes of education: *Group 1—No education or incomplete first year; Group 2—Complete first year or incomplete primary education; Group 3—Complete primary education or incomplete secondary education; Group 4—Complete secondary education or incomplete undergraduate; Group 5—Complete undergraduate; and Group 6—Graduate studies.*

The panel is unbalanced, and on average, each household is interviewed 7.81 times (months) per year. We possess microdata at the individual level for this survey, beginning in January 2006 through to May 2015 ($T = 113$ months). Overall, our sample has $N \times T = 164{,}479$ responses. Decomposing our sample into N and T gives the following breakdown: $T = 113$ months, and an average of approximately $N = 1456$ individuals per month. The key question of interest for us is the following: *"In your opinion, how much will be Brazilian inflation over the next 12 months?"* We excluded outliers in the data set whenever the respondent answered that inflation would be greater than 100 % in the next 12 months. In our sample, it was rare to observe 12-month inflation greater than 10 %, so we considered this type of response completely out of scope.[5]

3.1.2 Focus survey of the Central Bank of Brazil

The *Focus Survey* of forecasts of the Central Bank of Brazil (BCB) contains daily (working days) forecasts from roughly 250 registered institutions since 1999, the year when Brazil implemented its *Inflation-Targeting Regime*.[6] About 100 of these institutions are actively feeding the database with forecasts on any given day. Institutions include professional forecasters—commercial banks, asset management firms,

---

[4] See consumer survey: methodological features (April 2013), available at http://portalibre.fgv.br.

[5] With this criterium, we excluded 349 observations of a total of 164,479.

[6] The *Focus survey* is widely used in Brazil, and its excellence has been internationally recognized. In 2010, the survey received the Certificate of Innovation Statistics, due to a second place in the II Regional Award for Innovation in Statistics in Latin America and Caribbean, offered by the World Bank.

consulting firms, non-financial firms or institutions, academics, etc. Participants can provide forecasts for a large number of economic variables, e.g., inflation using different price indices, interest and exchange rates, GDP, industrial production, balance of payments accounts, and fiscal variables and for different forecast horizons.

Our focus here is on Brazilian inflation, measured by the Brazilian Consumer Price Index—IPCA, which is the official inflation target of the Brazilian *Inflation-Targeting Regime*. Our sample covers daily inflation forecasts collected from January 2nd, 2004 until May, 28th, 2015 (2861 workdays).[7] In each day t, $t = 1, \ldots, T$, survey respondent i, $i = 1, \ldots, N$, may inform her/his expectations regarding inflation rates all the way up to the next 18 months, as well as for the next twelve months and for the next 5 years on a year-end basis. The data set regarding the twelve-month-ahead inflation forecasts forms an unbalanced panel ($N \times T$) containing an amount of 244,043 observations. Decomposing our total number of observations into N and T gives the following breakdown: $T = 2861$ daily observations and an average of $N = 85.3$ forecasters in our sample. For more information on this data set, see Carvalho and Minella (2012) and Marques (2013).

Despite the fact that the Focus Survey has a longer time span and a higher frequency than FGV's Consumer Survey (working day vs. monthly), our sample size is, most of the time, constrained by the span and frequency of the latter.

3.1.3 Consensus forecasts

Our target variable y_t—is Brazilian inflation—as measured by the Broad Consumer Price Index (IPCA) collected at the monthly frequency. The consumer forecasts $f_{i,t}^h$ from the FGV survey regarding expected inflation rate over the next 12 months ($h = 12$ months) are cross-sectionally averaged, forming the so-called consensus forecasts $\overline{f_{.,t}^h} = \frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h$. Besides the average over the entire data set (approximately 1500 households), labeled hereafter as "cons_all", the individual forecasts are also averaged by selected demographic groups, according to the education level of the survey participant or to its level of income. The breakdown by education yields the following consensuses: cons_educ_1_3; cons_educ_4 and cons_educ_5_6, respectively, educational levels groups 1 through 3, level 4, and group level 5 through 6. For income, we breakdown the groups numbered from 1 to 4, respectively: cons_inc_1; cons_inc_2; cons_inc_3; cons_inc_4. Figure 1 plots these eight time series of aggregate (cross-sectional mean) forecasts.

There is considerable heterogeneity across groups, either from an education or income point of view. Low-income households usually forecast higher inflation rates *vis-à-vis* the forecasts of higher-income families. Obviously, inflation is not perceived equally by all families, each with its own consumption bundle. Low-income households spend more on food, while higher-income households spend a larger proportion of their budget on housing, education, and leisure. By construction, the IPCA index

---

[7] Since the survey had a small cross-sectional coverage (small N) in its first years, we only considered *post*-2003 data.
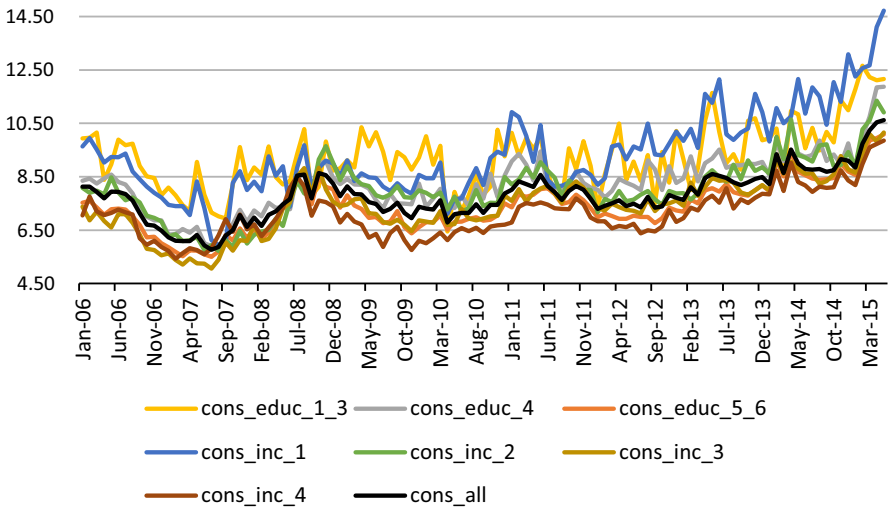
Fig. 1 Consensus inflation forecasts $\overline{f_{.,t}^h}$ from the FGV consumer survey (% 12-months-ahead)
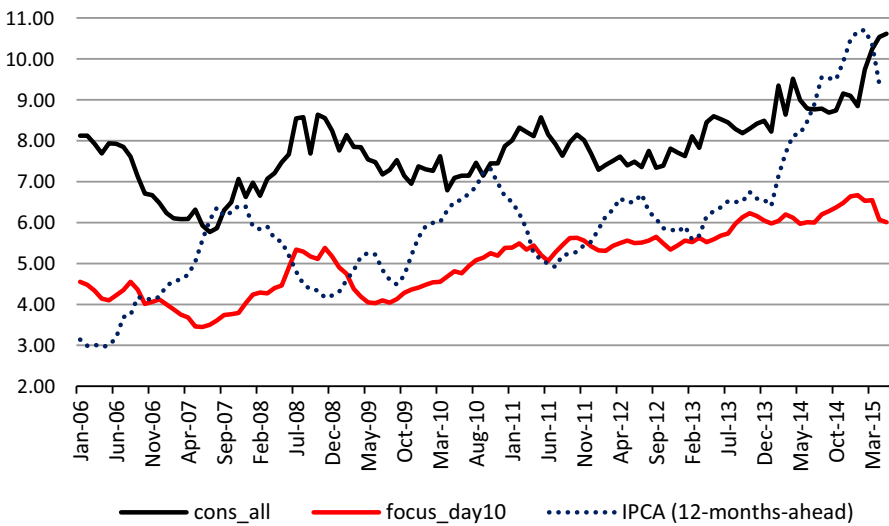
Fig. 2 Comparison of the consumer and market inflation forecasts $\overline{f_{.,t}^h}$ with the actual inflation rate y_t (% 12 months)

follows closely the cost-of-living of households with income well above the income level of our categories 1 and 2.

Figure 2 compares the consensuses of consumer forecasts (cons_all), professional (or market forecasts from the *Focus Survey*), and the twelve-month-ahead IPCA inflation rate. The latter is exactly what these forecasts are supposed to track. Market forecasts are collected at the 10th and 15th day of each month (focus_day10, focus_day15). This makes the focus consensus consistent with the days in which the

data are collected for the FGV consumer survey. On average, the consumer consensus has a positive bias, while the consensus of professional forecasters has a negative bias. It is also worth noting that the behavior of both consensuses is very different at the end of the sample period. While the consumer consensus follows the observed increase in inflation, the consensus of professional forecasters gets it completely wrong, predicting a decrease in inflation.

3.2 Empirical results

3.2.1 Descriptive statistics

Table 1 shows the pairwise sample correlations of the consensuses forecasts. Notice that all series are highly correlated. The consensus forecasts of consumers and of market agents show a sample correlation of 0.80.

Table 2 presents the sample mean and the respective estimation of the asymptotic standard error based on the number of sample observations $T = 113$. Next, we investigate whether or not the consensuses forecasts Granger Cause future inflation using Sims' test. More than half of the consensuses forecasts for consumers Granger Cause future inflation, i.e., are leading indicators of future inflation. This is also true for the consensus including all consumers. The same is true for both the consensuses of professional forecasters. Therefore, they will be useful as tools in forecasting future inflation.

3.2.2 Bias-correction devices

In Fig. 2, we provided soft evidence that the consensuses forecasts are biased as predictors of future inflation. Next, we formally test for zero bias employing the bias-corrected average forecast (BCAF) approach of Issler and Lima (2009). There, the estimate of the out-of-sample additive forecast bias $\overline{k^h}$ is:

$$\widehat{\overline{k^h}} = \frac{1}{NR} \sum_{t=T_1+1}^{T_2} \sum_{i=1}^{N} \left(y_t - f_{i,t}^h \right),$$ (9)

where its robust standard error is computed taking into account possible spatial and time dependence. Here, we employ the whole sample in estimating $\overline{k^h}$: $t = 1, 2, \ldots, T$, where the consensus data runs from January 2006 to July 2014, whereas the IPCA sample runs from December 2006 to June 2015.

Results of the zero-bias test are presented in Table 3. As expected from the plot in Fig. 2, all consumer forecasts showed a negative bias, i.e., consumers overestimated future inflation, while all professional forecasts underestimated future inflation. The formal evidence provided below calls for the use of bias-correction devices in constructing optimal forecasts.

A possible shortcoming of the test in Table 3 is that it is implicitly assumed that $\overline{\beta^h} = 1$. However, Gaglianone and Issler (2015) showed that, in general, there is

Table 1 Sample correlations

Variable	cons_all	cons_ed13	cons_ed4	cons_ed56	cons_inc1	cons_inc2	cons_inc3	cons_inc4	focus_d10	focus_d15
cons_all	1.000									
cons_ed13	0.768	1.000								
cons_ed4	0.950	0.694	1.000							
cons_ed56	0.986	0.702	0.904	1.000						
cons_inc1	0.874	0.808	0.900	0.822	1.000					
cons_inc2	0.945	0.708	0.911	0.930	0.785	1.000				
cons_inc3	0.957	0.667	0.909	0.965	0.802	0.892	1.000			
cons_inc4	0.927	0.662	0.832	0.950	0.785	0.812	0.890	1.000		
focus_d10	0.803	0.543	0.808	0.816	0.779	0.771	0.867	0.753	1.000	
focus_d15	0.803	0.542	0.806	0.819	0.776	0.768	0.869	0.758	0.999	1.000

Full sample: January 2006 to May 2015 ($T = 113$ observations)

Table 2 Mean forecast, SE, and Granger Causality test

Variable	Mean	SE	GC test
cons_all	7.78	0.09	0.0296**
cons_educ_1_3	9.21	0.11	0.5110
cons_educ_4	8.23	0.10	0.0219**
cons_educ_5_6	7.38	0.09	0.0423**
cons_inc_1	9.32	0.15	0.0175**
cons_inc_2	8.01	0.10	0.1171
cons_inc_3	7.41	0.10	0.1257
cons_inc_4	7.10	0.08	0.2206
focus_day10	5.05	0.08	0.0306**
focus_day15	5.06	0.08	0.0211**

Full sample: January 2006 to May 2015 ($T = 113$ observations)

Confidence interval based on asymptotic values: SE = sample SD/$\sqrt{T}$

The last column shows the p values of the Granger Causality test (12 lags) based on the null hypothesis: Forecast does not GC IPCA

** Rejection at a 5 % significance level

Table 3 BCAF additive bias $\overline{k^h}$ test (Issler and Lima 2009)

Variable	Mean	Bias	Robust SE	Zero-bias test (p value)
cons_all	7.62	−2.01	0.26	0.000
cons_educ_1_3	9.01	−3.41	0.27	0.000
cons_educ_4	8.06	−2.45	0.26	0.000
cons_educ_5_6	7.21	−1.61	0.26	0.000
cons_inc_1	9.02	−3.41	0.26	0.000
cons_inc_2	7.83	−2.22	0.28	0.000
cons_inc_3	7.24	−1.64	0.26	0.000
cons_inc_4	6.93	−1.33	0.25	0.000
focus_day10	4.92	0.68	0.21	0.001
focus_day15	4.93	0.68	0.21	0.001

Consensus sample: January 2006 to July 2014. IPCA sample: December 2006 to June 2015 (103 observations). Average IPCA (in 12 months) = 5.61 %. Unconditional (additive) bias is defined as: IPCA—mean (forecast). The zero-bias test is based on the null hypothesis H_0: $\overline{k^h} = 0$

Robust SE employs Newey–West. See Issler and Lima (2009)

an intercept and a slope bias term for the consensus forecast, suggesting the use of the Extended BCAF setup, where the parameters $\overline{k^h}$ and $\overline{\beta^h}$ are estimated by the generalized method of moments (GMM) based on y_t and $f_{i,t}^h$.[8]

Results of GMM estimation and of the no-bias test are presented in Table 4. We use a set of instruments z_{t-s} containing two lags of the consensus forecasts $\overline{f_{\cdot,t}^h}$, up

---

[8] The "iterative" procedure of Hansen et al. (1996) is employed in the GMM estimation, and the initial weight matrix is the identity.

Table 4 Extended BCAF estimation and No Bias test

Variable	$\widehat{\overline{k^h}}$	Robust SE	t stat p value	$\widehat{\overline{\beta^h}}$	Robust SE	t stat p value	OIR test p value	No Bias test p value
cons_all	−0.11	2.59	0.97	1.34	0.46	0.00	0.115	0.000
cons_educ_1_3	2.77	3.40	0.41	1.08	0.59	0.07	0.207	0.000
cons_educ_4	0.65	2.12	0.76	1.30	0.37	0.00	0.103	0.000
cons_educ_5_6	−1.64	2.77	0.55	1.54	0.49	0.00	0.197	0.000
cons_inc_1	0.33	1.54	0.83	1.54	0.27	0.00	0.128	0.000
cons_inc_2	0.03	2.60	0.99	1.37	0.46	0.00	0.060	0.000
cons_inc_3	−2.12	2.65	0.42	1.64	0.47	0.00	0.170	0.000
cons_inc_4	−1.47	2.75	0.59	1.44	0.49	0.00	0.268	0.000
focus_day10	−2.28	1.29	0.08	1.26	0.22	0.00	0.704	0.000
focus_day15	−2.24	1.26	0.08	1.25	0.22	0.00	0.702	0.000

Consensus sample: January 2006 to July 2014. IPCA sample: December 2006 to June 2015 (103 observations). Robust standard errors (SE) from a GMM estimation. The column "OIR test" denotes the p values from the Over-Identifying restriction (OIR) J test due to Hansen (1982)

The column "No-Bias test" shows the p values based on the null hypothesis $H_0: [\overline{k^h} = 0; \ \overline{\beta^h} = 1]$

Table 5 Forecast combination (Granger and Ramanathan 1984) $IPCA_{t+12} = \alpha + \beta_1 \text{cons\_all}_t + \beta_2 \text{focus\_day10}_t + \varepsilon_{t+12}$

Parameter	Unrestricted regression		Restricted regression	
	Full sample	Last 5 years	Full sample	Last 5 years
$\widehat{\alpha}$	4.723**	1.032	2.578***	0.646
	(2.192)	(2.627)	(0.851)	(0.769)
$\widehat{\beta_1}$	−0.865**	0.064	−0.703**	0.084
	(0.378)	(0.396)	(0.326)	(0.316)
$\widehat{\beta_2}$	1.518***	0.873***	–	–
	(0.336)	(0.277)		
Adjusted R^2	0.385	0.279	0.348	0.289

Consensus sample: January 2006 to July 2014. IPCA sample: December 2006 to June 2015
Robust (HAC) standard errors in parentheses. Restricted regression imposes $\beta_1 + \beta_2 = 1$
*, **, and *** indicate, respectively, significance at 10, 5, and 1 % level

to four lags of the output gap[9] and two lags of the Commodity Price Index (IC-Br), which weighting structure is designed to measure the impact of commodity prices on Brazilian consumer inflation[10]; see Appendix for results using an alternative set of instruments. Hansen's Over-Identifying Restriction (OIR) test is employed in order to check the validity of GMM estimates. At the usual significance levels, we do not observe any rejection for the OIR test. The no-bias test consists of a joint test of the null hypothesis $H_0: [\overline{k^h} = 0; \overline{\beta^h} = 1]$. We reject the null of no-bias on all occasions, showing that the consensus forecasts are a bias-ridden version of the conditional expectation $\mathbb{E}_{t-h}(y_t)$.

From Eq. (5), note that, under the null $H_0: [\overline{k^h} = 0; \overline{\beta^h} = 1]$, the aggregate forecast $\overline{f_{.,t}^h}$ should converge in probability to the conditional expectation $\mathbb{E}_{t-h}(y_t)$, with no need for bias correction. However, if the null is rejected, we should expect a corrected version of the consensus forecast to produce superior forecasts *vis-à-vis* the consensus forecasts. These are exactly our findings below.

3.2.3 Forecast combination of consumer and market expectations

Next, we apply the Granger and Ramanathan (1984) forecast combination technique for the entire sample. It consists of a standard OLS regression (including an intercept) of IPCA inflation on the two consensus forecasts—consumer and market, respectively: cons_all and focus_day10. We run those regressions in two different ways: When the two slope coefficients are unrestricted and when they are restricted to add up to unity. Also, we cover two different sample periods: the whole sample and only the last five years. Results are presented in Table 5. Notice that, for the last 5 years, the consumer consensus is not significant, which does not happen for the whole sample.

---

[9] The Hodrick-Prescott (HP) filtered output gap is computed from the log real monthly GDP (IBRE/FGV).

[10] The adopted set of instruments is slightly modified for the lower education groups (cons_educ_1_3) and the market forecasts (focus_day10; focus_day15).

The estimated slopes can be interpreted as optimal weights under a MSE loss function. The combined (consumer-market) forecast is given by:

$$\widetilde{fcm}_t^h = \widehat{\alpha} + \widehat{\beta}_1 \mathrm{cons\_all}_t + \widehat{\beta}_2 \mathrm{focus\_day}10_t.$$

3.2.4 Out-of-sample forecasting exercise

In this section, we conduct a truly out-of-sample forecasting exercise, incorporating what we have learned from the previous sections. We compare different forecast methods by computing their out-of-sample mean-squared error (MSE) of forecasting. Our whole sample for the experiment runs from January 2006 through July 2014 for the forecasts and from February, 2006 through June 2015 for the realizations of inflation, comprising 103 and 113 observations, respectively. We leave the last 60 observations (5 years, comprising July 2010 through June 2015) for a pseudo out-of-sample evaluation of several forecasting strategies (or methods), which are estimated with a *growing* window. Thus, model parameters are estimated recursively as we incorporate every new time-series observation, one at a time. These forecasting methods include: a "best" model, chosen by information criteria, a Phillips-curve based forecast, the Granger and Ramanathan (1984) forecast-combination technique, the Capistrán and Timmermann (2009) consensus regression, the average forecast (or consensus forecast), the bias-corrected average forecast (BCAF), proposed by Issler and Lima (2009), and the extended BCAF, proposed by Gaglianone and Issler (2015). For the last four strategies, we also present the MSEs by category of income and education.

The best model using information criteria is the AR(1) model for the monthly inflation rate, see Cordeiro et al. (2015) and Gaglianone et al. (2016). The AR(1) model for the monthly inflation rate y_t is as follows:

$$y_t = c + \beta y_{t-1} + \varepsilon_t. \tag{10}$$

From a sample with $t = 1, \ldots, T$ observations, the respective estimates $\left[\widehat{c}_T ; \widehat{\beta}_T\right]'$ are computed. The respective h-step-ahead forecast ($h > 1$) is given by

$$\widehat{y_{T+h}} = \left(\widehat{\beta}_T\right)^h y_T + \sum_{i=1}^{h} \widehat{c}_T \left(\widehat{\beta}_T\right)^{i-1}, \tag{11}$$

and the cumulative twelve-month rate ahead $y_{T+12}^{12m\_rate}$, in percentage is:

$$\widehat{y_{T+12}^{12m\_rate}} = 100 \left(\left(\prod_{h=1}^{12} \left(1 + \frac{\widehat{y_{T+h}}}{100} \right) \right) - 1 \right). \tag{12}$$

We also estimate a Phillips-curve model (PC), which has a long tradition in forecasting inflation (for a comprehensive survey, see Stock and Watson 2009). The Brazilian Central Bank (BCB) has had several years of experience fitting different versions of the Phillips curve, and we benefit from that in our forecasting exercise. Here, we consider

a hybrid (New Keynesian) version of the Phillips curve, which includes backward and forward looking terms, lagged imported inflation (to consider a pass-through channel), and a lagged output gap. In the BCB experience, what works best for forecasting IPCA inflation is to disaggregate inflation into two main components: market prices (roughly 75 % of IPCA) and the so-called regulated and monitored prices (approximately 25 % of IPCA), which contains consumer items that are relatively insensitive to domestic demand and supply conditions or that are in some way regulated by a public agency. Since these two groups have completely different time dynamics, they are modeled separately. We first estimate a Phillips curve for inflation of the market prices basket, as follows:

$$\pi_t^{\text{market}} = \alpha_0 + \alpha_1 \pi_{t-1}^{\text{market}} + \alpha_2 \pi_{t|t-1}^{\text{exp}} + \alpha_3 \pi_{t-4}^{\text{imp}} + \alpha_4 g_{t-1} + \varepsilon_t, \qquad (13)$$

where π_t^{market} is the monthly inflation of market prices, $\pi_{t|t-1}^{\text{exp}}$ denotes the one-month ahead expected inflation (from the Focus survey), π_t^{imp} is what has been labeled as *imported inflation* (R\$/US\$ exchange rate variation plus the U.S. PPI), and g_t is the output gap.[11] We also impose the coefficient restriction: $\alpha_1 + \alpha_2 + \alpha_3 = 1$, to guarantee a vertical long-run Phillips curve. For the so-called *monitored prices*, we estimate an auxiliary AR model. The usual criteria for lag order selection indicate one lag only (and the estimated autoregressive coefficient is 0.53). Then, we use each model to produce h-step ahead point forecasts ($h = 1, \ldots, 12$ months) which are, then, aggregated through time to form a 12-month cumulated forecast for each component.[12] Both forecasts are then aggregated (using corresponding CPI weights) in order to generate an aggregated 12-month ahead point forecast for the whole IPCA inflation.

Table 6 presents the comparison of the out-of-sample MSE of the different forecast strategies listed above. First, the MSE of the AR(1) model is higher when compared to that of the Granger and Ramanathan combination of the consensuses of consumers and of professional forecasters, and it is lower than that of the average forecasts—exceptions are the consensus for the highly educated forecasters, higher income forecasters, and professional forecasters (Focus Survey). Second, when we compare the MSE of the average forecast, BCAF and Extended BCAF, there is a clear *pecking order*: The extended BCAF is more accurate than the BCAF, which, in turn, dominates the average forecast. This validates the view that the bias corrections performed either by the BCAF, or by the extended BCAF, are a useful device for forecasting using surveys. Third, when we compare the best forecasts for all consumers (Extended BCAF, cons_all) with the best forecasts for all professional forecasters (Extended BCAF, focus_day15), we observe a reduction in MSE in favor of profes-

---

[11] The output gap is based on the seasonally adjusted IBC-BR index of economic activity. The Hodrick–Prescott (HP) filter is employed to generate the output gap in a recursive estimation scheme, that is, we re-construct the entire output gap series for each new observation added to the estimation sample along the out-of-sample exercise (and then re-estimate the Phillips curve to construct new h-step ahead forecasts). This way, we perform a "pseudo" real-time PC-forecast (based on the last available vintage of the IBC-BR).

[12] The h-step ahead forecasts are constructed using an iterated procedure, which is trivial in the case of the AR model. In the case of the PC and $h > 1$, we use the previous PC-forecast (from $h - 1$), random walk forecasts for the expected and imported inflation rates and the AR(1) forecast for the output gap (autoregressive coefficient of 0.95, based on in-sample estimations).

sional forecasters of 22.5 %, which is sizable. This happens despite the fact that, for consumers, we are averaging the forecast of approximately 1500 individuals, whereas for professional forecasters, we employ about 85 individuals. This may be a sign that it matters to employ informed and well-trained professionals in forecasting inflation—a relief for the profession as a whole.

Forecasts using the Phillips curve ranked relatively well here—a median result— better than the AR(1) model, the average forecast, and the Capistrán and Timmermann method, but worse than the Granger–Ramanathan combination, the BCAF, and the extended BCAF. The Capistrán and Timmermann (2009) consensus regression did slightly worse than the Phillips curve. A direct comparison of Capistrán and Timmermann's method with the extended BCAF is interesting here, since, in principle, both could potentially identify the *intercept* and *slope* biases present in the consensus forecast, respectively, denoted by $\overline{k^h}$ and $\overline{\beta^h}$ in Sect. 2. Indeed, the MSEs of the Extended BCAF and that of the Capistrán and Timmermann method are very different across all levels of education and income for consumers, as well as for all days in the Focus Survey. When we consider all consumers, the MSE of the extended BCAF is 0.42 of that of Capistrán and Timmermann's method, a sizable reduction.

In order to check whether the differences in MSE listed in Table 6 are significant, we employ the equal predictive accuracy test of Clark and West (2007) for nested models, and the equal predictive ability test of Diebold and Mariano (1995) for non-nested models. All MSE comparisons are direct between the MSE of the extended BCAF and that of any specific cell in Table 6. In most cases, the extended BCAF produces an MSE that is statistically different (and smaller) than that of the direct competitor. Details are as follows: The average forecast and the BCAF have an inferior out-of-sample MSE *vis-à-vis* the extended BCAF forecast. Comparisons of the extended BCAF with the AR(1) model using a Diebold–Mariano test for equal variances shows that the former is statistically superior to the latter (at the 10 % significance level) in almost all cases. The Extended BCAF for the market forecasts also statistically improves (at a 10 % level) the out-of-sample accuracy compared to the Granger and Ramanathan (1984) forecast combination approach. Also, it statistically improves on the Capistrán and Timmermann (2009) consensus-forecasting approach almost everywhere.

All in all, if we had to suggest a single forecasting strategy for Brazilian inflation one-year ahead, we would suggest the use of extended BCAF based on the Focus Survey consensus (15th working day).

3.2.5 Rationality tests

In testing for rationality, we employed two alternative approaches. The first is based on the well-known Mincer–Zarnowitz regression. It is only valid if one assumes that the forecaster minimizes the MSE risk function. For other functions, the test no longer applies,[13] and one must find alternative ways of testing rationality. Here,

---

[13] For example, an asymmetric loss function (e.g., lin-lin) could reflect the forecaster's different weighting scheme of positive forecast errors *vis-à-vis* the negative ones. In this setup, a non-trivial forecast bias may arise, for instance, reflecting the asymmetric loss function (instead of lack of rationality). For further

Table 6 Forecast evaluation: mean-squared error (MSE) out of sample

Variable	AR(1)	PC	GR	CT	Average forecast	BCAF	Extended BCAF
cons_all	2.344*	1.240	0.921	1.795*	3.406*	1.026	0.760
	[0.070]	[0.274]	[0.311]	(0.000)	(0.000)	(0.129)	
cons_educ_1_3	2.344*	1.240	0.921	1.757*	10.148*	1.927*	1.417
	[0.074]	[0.638]	[0.892]	(0.000)	(0.000)	(0.017)	
cons_educ_4	2.344*	1.240	0.921	1.789*	5.658*	0.863*	0.745
	[0.056]	[0.232]	[0.257]	(0.000)	(0.000)	(0.092)	
cons_educ_5_6	2.344*	1.240	0.921	1.689*	2.287*	1.047	0.758
	[0.066]	[0.264]	[0.315]	(0.000)	(0.008)	(0.196)	
cons_inc_1	2.344	1.240	0.921	1.568*	11.845*	1.149*	0.740
	[0.128]	[0.200]	[0.345]	(0.000)	(0.000)	(0.001)	
cons_inc_2	2.344*	1.240	0.921	1.868*	4.650*	0.939	0.783
	[0.047]	[0.255]	[0.354]	(0.000)	(0.000)	(0.103)	
cons_inc_3	2.344*	1.240	0.921	1.539*	2.784*	0.878	0.768
	[0.059]	[0.288]	[0.308]	(0.000)	(0.001)	(0.251)	
cons_inc_4	2.344*	1.240	0.921	1.618*	1.523*	1.308	0.787
	[0.086]	[0.284]	[0.344]	(0.000)	(0.050)	(0.174)	
focus_day10	2.344*	1.240	0.921*	1.104	1.295	0.624	0.590
	[0.050]	[0.209]	[0.098]	(0.248)	(0.289)	(0.454)	
focus_day15	2.344*	1.240	0.921*	1.075	1.282	0.624	0.589
	[0.052]	[0.205]	[0.099]	(0.247)	(0.296)	(0.443)	

Forecast evaluation sample = last 60 observations—July 2010 through June 2015. PC denotes the Phillips curve forecast. GR means the combined forecast of Granger and Ramanathan (1984), based on an OLS regression with intercept and "cons_all" and "focus_day10" forecasts as regressors. CT denotes the forecast from the OLS regression of Capistrán and Timmermann (2009) of inflation on the consensus and an intercept. The second, third, and fourth columns show [in brackets] the *p* values of the equal predictive ability test of Diebold–Mariano (1995) between the AR(1), PC, and GR, respectively, compared to the Extended BCAF (on each row). The fifth, sixth, and seventh columns show (in parenthesis) the *p* values of the equal predictive accuracy test of Clark and West (2007), which compares the CT forecast, the average forecast and the BCAF, respectively, with the Extended BCAF (on each row). In all cases, * indicates a rejection of the null at a 10 % level

we employed the quantile Mincer–Zarnowitz regression test proposed by Gaglianone et al. (2011).

Let y_t be the 12-month inflation rate measured by IPCA, $f_{i,t}^h$ be the respective inflation forecast of survey participant i formed at period $t-h$, and $\overline{f}_{.,t}^h$ be the consensus forecast. The cross-sectional dimension is considered in two ways: (i) disaggregated data, with an individual OLS regression for each $i = 1, \ldots, N$[14]; and (ii) aggregated

---

Footnote 13 continued

details, see Batchelor (2007) which presents several arguments consistent with forecasters having a nonquadratic loss function.

[14] In each test, with disaggregated data (individual OLS regressions), we only considered survey participants with more than 20 available forecasts.

Table 7 Rationality (MZ) tests for consumers

Variable	Individual forecasts (% of rational)	Consensus forecasts (MZ test, p value)
cons_all	35	0.000
cons_educ_1_3	26	0.000
cons_educ_4	31	0.000
cons_educ_5_6	40	0.000
cons_inc_1	22	0.000
cons_inc_2	31	0.000
cons_inc_3	35	0.000
cons_inc_4	37	0.000

In the second column, the % of rational forecasters is based on p value >0.05 in the MZ individual rationality test and on consumers with at least 20 observations in time dimension
Consensus sample: January 2006 to July 2014. IPCA sample: December 2006 to June 2015 (103 observations). Robust standard errors (SE) from a Newey–West covariance matrix

data (consensus forecast), with OLS regression. We focus on the widely employed rationality test of Mincer and Zarnowitz (1969), hereafter MZ. First, we consider disaggregate data, with an individual OLS regression for each $i = 1, 2, \ldots, N$. The individual-forecasts OLS regressions are given as follows:

$$y_t = \alpha_i + \beta_i f_{i,t}^h + \varepsilon_{i,t}, \tag{14}$$

where y_t denotes observed inflation on period t, and $f_{i,t}^h$ denotes individual i forecast of y_t using only information up to period $t - h$. The MZ test is based on a Wald test of the joint null hypothesis of rationality, H_0: $[\alpha_i = 0; \beta_i = 1]$ for each i. We then compute the proportion of agents for which we do not reject the null of rationality. Next, we consider aggregate data (consensus forecast), with OLS regression in time-series data. The consensus-forecast OLS regression is given by:

$$y_t = \alpha + \beta \overline{f_{\cdot,t}^h} + \varepsilon_t, \tag{15}$$

where the rationality test is based on a Wald test of the joint null hypothesis H_0: $[\alpha = 0; \beta = 1]$. In testing all the null hypotheses listed above, we have employed robust (HAC) standard errors. Results for consumer forecasts are summarized in Table 7.[15]

When we consider tests using Mincer–Zarnowitz regressions with disaggregate data, individual OLS regressions for each i show that 35 % of the consumers indeed pass rationality tests at the 5 % significance level. Moreover, this percentage increases to 40 % and 37 % for the consensuses of higher educated consumers and for those with higher income, respectively. However, when we analyze test results using the

[15] We leave the rationality tests for market forecasters for future research, since the focus here is on leveraging the consumer data for forecasting purposes.

consensus forecasts with OLS regressions, the null is always rejected at the 5 % level, which is consistent with results previously found in the literature, where rationality was rejected overwhelmingly in Mincer–Zarnowitz tests (Coibion and Gorodnichenko 2012), and the results are shown in Tables 3 and 4, in which the null of zero bias is strongly rejected.[16]

Next, we test for rationality by employing the quantile Mincer–Zarnowitz regression test proposed by Gaglianone et al. (2011). Similarly to the OLS-MZ regression, the idea is to estimate a quantile regression (QR), for a given quantile level τ, of the form: $Q_\tau(y_t | \mathcal{F}_{i,t-h}) = \theta_{0,i}(\tau) + \theta_{1,i}(\tau) f_{i,t}^h$, in the case of individual forecasts, and of the form: $Q_\tau(y_t | \mathcal{F}_{t-h}) = \theta_0(\tau) + \theta_1(\tau) \overline{f_{.,t}^h}$ for the consensus forecast.

The QR-MZ test is performed by using a Wald test of the joint null hypothesis $H_0 : [\theta_{0,i}(\tau) = 0; \theta_{1,i}(\tau) = 1]$ at a given level $\tau \in (0, 1)$, for individual forecasts, or $H_0 : [\theta_0(\tau) = 0; \theta_1(\tau) = 1]$ for the consensus forecast. The results are presented in Table 10 of the Appendix. Overall, there is an increase in the percentage of rational forecasters as long as the quantile level τ increases, suggesting that consumer's rationality in Brazil is more likely to occur at the right tail of the inflation conditional distribution.[17] This result is perfectly in line with the fact that consumers systematically overestimate inflation (which might be a sign that consumers assign more weight to positive forecast errors than to negative ones). With respect to the consensus forecast, similar to the result from the OLS-MZ test, we again strongly reject the null hypothesis of rationality (at a 1 % significance level) in the QR-MZ sense, for all considered cases.

We can summarize our results in this section as follows: We provide evidence that rationality test results based on a time series of individual forecasts are very different from those using the time series of the aggregate measure represented by the consensus forecast. When testing individuals separately, we find evidence of rationality for a subset of consumers, whereas we find the opposite for the consensus forecasts. These results are stronger when we focus on the mean forecast (standard MZ test), but are still present when we employ the quantile Mincer–Zarnowitz regression test proposed by Gaglianone et al. (2011). One possible explanation is that the group of "non-rational" consumers contaminates the result of the consensus.

4 Conclusions

In this paper, we investigate empirically whether or not combining forecasts from surveys of expectations is a helpful strategy for forecasting one-year-ahead CPI inflation for Brazil. Combining forecasts has been a promising strategy at least since the seminal paper of Bates and Granger (1969). The techniques used in this paper have their roots in the pioneering work of Davies and Lahiri (1995) on what could be labeled *a*

---

[16] Results in Table 7 reveal that there might be some useful information (for forecasting purposes) embodied in the consumer consensus forecast. Also, in forming the consensuses, one might consider, for instance, taking into account only those rational consumers when forming a "filtered-consensus" forecast, avoiding mixing information from the non-rational forecasters. We leave this for future research.

[17] Regarding income level 1, the % of rational consumers jumps from 30 % (at $\tau = 0.3$) to 62 % (at $\tau = 0.7$).

panel-data approach to forecasting. In this literature, some of the studies focused on forecast combination and others on density-forecast estimation, e.g., Issler and Lima (2009), Gaglianone et al. (2011), Gaglianone and Lima (2012, 2014), Gaglianone and Issler (2015), and Lahiri et al. (2015); see Baltagi (2013) for a broader discussion on this topic.

We investigate inflation forecasts from the *Brazilian Consumer Survey* conducted by FGV–IBRE, which consists of monthly quantitative and qualitative information from consumers in seven major capitals of the country, with a representative sample of the population and the rate of consumption of each capital with approximately 2000 consumers. Our sample ranges from January 2006 up to May 2015, forming an unbalanced panel with $NT = 164,479$ responses ($T = 113$ months; and an average of $N = 1456$ individuals per month). We also use The Focus Survey of the Central Bank of Brazil. It contains daily (working days) forecasts from roughly 250 registered institutions since 1999, the year when Brazil implemented its *Inflation-Targeting Regime*. About 100 of these institutions are actively feeding the database with forecasts on any given day. Institutions include professional forecasters working in commercial banks, asset management firms, consulting firms, non-financial firms or institutions, academics, etc.

Natural candidates to win a forecasting competition in the literature of surveys of expectations are the so-called consensus forecasts, i.e., the cross-sectional average among survey respondents. In an exploratory investigation, we first show that these consensus forecasts are a bias-ridden version of the conditional expectation of one-year ahead inflation. No-bias tests are conducted for the intercept and slope using the methods in Issler and Lima (2009) and Gaglianone and Issler (2015). The bias-test results reveal interesting features of the data: Consumers systematically overestimate inflation by 2.01 p.p., on average, whereas professional forecasters underestimate it by 0.68 p.p., on average, over the same sample. Furthermore, we show that these biases lead to rejection in the Mincer–Zarnowitz tests of rationality using the consensus forecasts for consumers, although from 22 to 40 % of consumers pass rationality tests at the individual level. We also apply the quantile Mincer–Zarnowitz regression test proposed by Gaglianone et al. (2011), where we found an increase in the percentage of rational forecasters as long as the quantile level τ increases, suggesting that consumer's rationality in Brazil is more likely to occur at the right tail of the inflation conditional distribution.

In a pseudo out-of-sample analysis, we evaluate seven different forecasting methods consisting of: A "best" model chosen by information criteria, the Granger and Ramanathan (1984) forecast combination technique, the Phillips-curve based forecasts, the Capistrán and Timmermann (2009) consensus-forecast method, the consensus forecast, the bias-corrected average forecast (BCAF), and the extended BCAF; we found very interesting results. The out-of-sample MSE of the AR(1) model is higher to that of the Granger and Ramanathan combination of the consensuses of consumers and of professional forecasters. It is also lower than the out-of-sample MSE of the average forecasts—with a few exceptions. When we compare the MSE of the average forecast, BCAF, and Extended BCAF, there is a clear *pecking order*: The extended BCAF is more accurate than the BCAF, which, in turn, dominates the average forecast. This validates the view that the bias corrections performed either by the BCAF, or by

the extended BCAF, are a useful device for forecasting using surveys. The results for the Capistrán and Timmermann (2009) consensus-forecast method are mixed. It beats the average forecast and the AR(1) model, but it is beaten by all the other combination methods and also by Phillips-curve based forecasts. The latter achieved a median result, since it was beaten by the Granger and Ramanathan combination, the BCAF, and the Extended BCAF.

Finally, when we compare the best forecasts for all consumers (Extended BCAF, cons_all) with the best forecasts for all professional forecasters (Extended BCAF, focus_day15), we observe a reduction in MSE in favor of professional forecasters of 22.5 %, which is sizable. This happens despite the fact that, for consumers, we are averaging the forecast of approximately 1500 individuals, whereas for professional forecasters, we employ a little less than 100 individuals. This may be a sign that it matters to employ informed and well-trained professionals in forecasting inflation—a relief for the profession.

All in all, if we had to suggest a single forecasting strategy for Brazilian inflation one-year ahead, we would suggest the use of extended BCAF based on the Focus Survey consensus (15th working day).

Appendix

In the Appendix, Tables 8–11, we provide complementary empirical evidence supporting some of the results presented in the text.

Table 8 Extended BCAF estimation

Variable	$\widehat{k^h}$	Robust SE	t stat p value	$\widehat{\beta^h}$	Robust SE	t stat p value	OIR test p value	No Bias test p value
cons_all	−1.04	2.91	0.72	1.49	0.51	0.00	0.121	0.000
cons_educ_1_3	1.53	4.90	0.76	1.29	0.85	0.13	0.277	0.000
cons_educ_4	−2.83	3.68	0.44	1.89	0.65	0.00	0.297	0.000
cons_educ_5_6	−3.47	3.76	0.36	1.85	0.66	0.01	0.378	0.000
cons_inc_1	−2.20	2.65	0.41	1.95	0.46	0.00	0.132	0.000
cons_inc_2	−14.58	13.76	0.29	3.90	2.42	0.11	0.961	0.000
cons_inc_3	−7.75	5.31	0.14	2.61	0.93	0.01	0.899	0.000
cons_inc_4	−2.07	3.11	0.51	1.54	0.55	0.00	0.270	0.000
focus_day10	−3.04	1.50	0.04	1.39	0.26	0.00	0.994	0.000
focus_day15	−2.95	1.45	0.04	1.38	0.25	0.00	0.989	0.000

GMM estimation based on the set of instruments: (lagged 1–5 months) aggregate forecasts

Consensus sample: January 2006 to July 2014. IPCA sample: December 2006 to June 2015 (103 observations). Robust standard errors (SE) from a GMM estimation of the BCAF

Set of instruments: lagged forecasts, from 1 to 5 months. The column "OIR test" denotes the p values from the over-identifying restriction (OIR) J test due to Hansen (1982)

The column "No-Bias test" shows the p values based on the null hypothesis $Ho: [\overline{k^h} = 0; \overline{\beta^h} = 1]$

Table 9 Consensus forecast OLS regression of Capistrán and Timmermann (2009): $y_t = a + b\overline{f^h_{\cdot,t}} + \epsilon_t$

Variable $f^h_{\cdot,t}$	$\hat{a}$	NW SE	t stat p value	$\hat{b}$	NW SE	t stat p value	Adjusted R^2	No Bias test p value
cons_all	3.36	2.26	0.14	0.30	0.31	0.34	0.0258	0.0000
cons_educ_1_3	4.51	1.54	0.00	0.12	0.18	0.50	0.0009	0.0000
cons_educ_4	2.95	1.84	0.11	0.33	0.24	0.16	0.0505	0.0000
cons_educ_5_6	3.11	2.03	0.13	0.35	0.29	0.23	0.0438	0.0000
cons_inc_1	2.45	1.48	0.10	0.35	0.16	0.04	0.1253	0.0000
cons_inc_2	3.90	1.89	0.04	0.22	0.25	0.38	0.0204	0.0000
cons_inc_3	2.76	1.67	0.10	0.39	0.23	0.09	0.0869	0.0000
cons_inc_4	3.11	2.04	0.13	0.36	0.31	0.24	0.0394	0.0000
focus_day10	1.49	1.46	0.31	0.84	0.28	0.00	0.2689	0.0048
focus_day15	1.44	1.45	0.32	0.85	0.28	0.00	0.2749	0.0052

Consensus sample: January 2006 to July 2014. IPCA sample: December 2006 to June 2015 (103 observations). Newey–West (1987)'s standard errors (SE) The column "No-Bias test" shows the p values based on the joint null hypothesis H_0: $[a = 0; b = 1]$

Table 10 Rationality (QR-MZ) tests for consumers

Variable	$\tau = 0.3$ (%)	$\tau = 0.4$ (%)	$\tau = 0.5$ (%)	$\tau = 0.6$ (%)	$\tau = 0.7$ (%)
cons_all	36	41	46	49	47
cons_educ_1_3	34	37	42	47	49
cons_educ_4	30	35	44	53	58
cons_educ_5_6	40	46	46	47	42
cons_inc_1	30	34	42	57	62
cons_inc_2	27	32	51	56	58
cons_inc_3	41	43	49	54	55
cons_inc_4	38	44	45	43	38

Table entries are the % of rational forecasters based on p value > 0.05 in the QR-MZ individual rationality test and on consumers with at least 20 observations in time dimension

Forecasts' sample: January 2006 to July 2014. IPCA sample: December 2006 to June 2015

Table 11 Phillips curve estimation $\pi_t^{market} = \alpha_0 + \alpha_1 \pi_{t-1}^{market} + \alpha_2 \pi_{t|t-1}^{exp} + \alpha_3 \pi_{t-4}^{imp} + \alpha_4 g_{t-1} + \varepsilon_t$

Parameter	Full sample
$\widehat{\alpha}_0$	0.0005 ** (0.0002)
$\widehat{\alpha}_1$	0.6054 *** (0.0914)
$\widehat{\alpha}_2$	0.3787 *** (0.0914)
$\widehat{\alpha}_3$	0.0160 ** (0.0074)
$\widehat{\alpha}_4$	0.0156 * (0.0081)
Adjusted R^2	0.394

Monthly frequency. Sample: January 2006 to June 2015. Robust (HAC) standard errors in parentheses

Restricted regression imposes: $\alpha_1 + \alpha_2 + \alpha_3 = 1$

*, **, and ***indicate, respectively, significance at 10, 5, and 1 % level

References

Ang A, Bekaert G, Wei M (2007) Do macro variables, asset markets or surveys forecast inflation better? J Monet Econ 54:1163–1212

Baltagi BH (2013) Panel Data Forecasting. In: Elliott G, Timmermann A (eds) Handbook of economic forecasting, vol 2. Elsevier, New York, pp 995–1024

Batchelor R (2007) Bias in macroeconomic forecasts. Int J Forecast 23:189–203

Bates JM, Granger CWJ (1969) The combination of forecasts. Oper Res Q 20:309–325

Capistrán C, Timmermann A (2009) Forecast combination with entry and exit of experts. J Bus Econ Stat 27:428–440

Carvalho FA, Minella A (2012) Survey forecasts in Brazil: a prismatic assessment of epidemiology, performance, and determinants. J Int Money Finance 31(6):1371–1391

Christoffersen PF, Diebold FX (1997) Optimal prediction under asymmetric loss. Econ Theory 13:808–817

Clark TE, West KD (2007) Approximately normal tests for equal predictive accuracy in nested models. J Econom 138:291–311

Coibion O, Gorodnichenko Y (2012) What can survey forecasts tell us about information rigidities? J Polit Econ 120:116–159

Cordeiro YAC, Gaglianone WP, Issler JV (2015) Inattention in individual expectations. Working paper no. 395, Central Bank of Brazil

Davies A, Lahiri K (1995) A new framework for analyzing survey forecasts using three-dimensional panel data. J Econom 68:205–227

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253–263

Elliott G, Komunjer I, Timmermann A (2008) Biases in macroeconomic forecasts: irrationality or asymmetric loss? J Eur Econ Assoc 6(1):122–157

Elliott G, Timmermann A (2004) Optimal forecast combinations under general loss functions and forecast error distributions. J Econom 122:47–79

Faust J, Wright JH (2013) Forecasting inflation. Handbook of economic forecasting, vol 2A, Chap 1, pp 3–56. Ed. Elsevier B.V

Gaglianone WP, Giacomini R, Issler JV, Skreta V (2016) Incentive-driven inattention. Getulio Vargas Foundation, Mimeo

Gaglianone WP, Issler JV (2015) Microfounded forecasting. Ensaios Econômicos EPGE no. 766, Getulio Vargas Foundation

Gaglianone WP, Lima LR (2012) Constructing density forecasts from quantile regressions. J Money Credit Bank 44(8):1589–1607

Gaglianone WP, Lima LR (2014) Constructing optimal density forecasts from point forecast combinations. J Appl Econom 29(5):736–757

Gaglianone WP, Lima LR, Linton O, Smith DR (2011) Evaluating value-at-risk models via quantile regression. J Bus Econ Stat 29(1):150–160

Genre V, Kenny G, Meyler A, Timmermann A (2013) Combining expert forecasts: can anything beat the simple average? Int J Forecast 29:108–121

Granger CWJ (1969) Prediction with a generalized cost of error function. Oper Res Q 20(2):199–207

Granger CWJ, Ramanathan R (1984) Improved methods of combining forecasting. J Forecast 3:197–204

Hansen LP (1982) Large sample properties of generalized method of moments estimators. Econometrica 50:1029–1054

Hansen LP, Heaton J, Yaron A (1996) Finite-sample properties of some alternative GMM estimators. J Bus Econ Stat 14:262–280

Issler JV, Lima LR (2009) A panel data approach to economic forecasting: the bias-corrected average forecast. J Econom 152(2):153–164

Lahiri K, Peng H, Sheng X (2015) Measuring uncertainty of a combined forecast and some tests for forecaster heterogeneity. CESifo working paper no. 5468

Marques ABC (2013) Central Bank of Brazil's market expectations system: a tool for monetary policy. Bank for International Settlements, vol 36 of IFC Bulletins, pp 304–324

Mincer JA, Zarnowitz V (1969) The evaluation of economic forecasts. In: Mincer Jacob (ed) Econmic forecasts and expectations. National Bureau of Economic Research, New York

Palm FC, Zellner A (1992) To combine or not to combine? Issues of combining forecasts. J Forecast 11(8):687–701

Patton AJ, Timmermann A (2007) Testing forecast optimality under unknown loss. J Am Stat Assoc 102(480):1172–1184

Phillips PCB, Moon HR (1999) Linear regression limit theory for nonstationary panel data. Econometrica 67:1057–1111

Stock J, Watson M (2002a) Macroeconomic forecasting using diffusion indexes. J Bus Econ Stat 20:147–162

Stock J, Watson M (2002b) Forecasting using principal components from a large number of predictors. J Am Stat Assoc 97(460):1167–1179

Stock J, Watson M (2006) Forecasting with many predictors. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of economic forecasting. North-Holland, Amsterdam, pp 515–554

Stock J, Watson M (2009) Phillips curve inflation forecasts. In: Fuhrer Jeffrey, Kodrzycki Yolanda, Little Jane, Olivei Giovanni (eds) Understanding inflation and the implications for monetary policy. MIT Press, Cambridge

Timmermann A (2006) Forecast combinations. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of economic forecasting. North-Holland, Amsterdam, pp 135–196

West K (1996) Asymptotic inference about predictive ability. Econometrica 64:1067–1084